

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:02:09 ; Search time 55.1415 Seconds

(without alignments)  
4469.091 Million cell updates/sec

Title: US-09-823-394-2

Perfect score: 6183

Sequence: 1 MKTFSSFFLVTTFEFSRF.....GFTTMDMSIKVEPSGKL 1196

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

7. number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6140	99.3	1196	10	O22476
2	3050.5	49.3	1121	10	O24276
3	2643.5	42.8	1192	10	O942F3
4	2643.5	42.6	1164	10	O9ARC8
5	2623.5	42.4	1166	10	O9LWF3
6	2582.5	41.8	1166	10	O9ZWC8
7	2509.5	40.6	1143	10	O9ARF3
8	2240	36.2	1110	10	O9ZPS9
9	1608.5	26.0	1110	10	O94LW2
10	1593.5	25.8	1294	10	O9LW8
11	1503	24.3	1420	10	O9XEL3
12	1419	23.0	1232	10	O9SNI3
13	1402.5	22.7	1036	10	O9FN37
14	1380.5	22.3	1102	10	O9LWPO
15	1380	22.3	1079	10	O9CA77
16	1370.5	22.2	1141	10	O9LHP4

17	1367	22.1	1133	10	O9SH12	O9sh12 arabidopsis
18	1364	22.1	1124	10	O49318	O49318 arabidopsis
19	1357.5	22.0	1003	10	O49345	O49345 arabidopsis
20	1353	21.9	1012	10	O9LW24	O9LW24 glycine max
21	1353	21.9	1236	10	O9F123	O9F123 arabidopsis
22	1349.5	21.8	1095	10	O9C755	O9C755 arabidopsis
23	1348	21.8	1008	10	O9LW26	O9LW26 arabidopsis
24	1346	21.8	1012	10	O9LW25	O9LW25 arabidopsis
25	1340.5	21.7	1134	10	O65510	O65510 glycine max
26	1339	21.7	1173	10	O9F128	O9F128 arabidopsis
27	1335	21.6	1064	10	O9FX19	O9FX19 arabidopsis
28	1332	21.4	1155	10	O8SB69	O8SB69 arabidopsis
29	1320	21.3	1140	10	O9LR04	O9LR04 arabidopsis
30	1313	21.2	1003	10	O9MOC7	O9MOC7 arabidopsis
31	1310	21.2	1002	10	O9M221	O9M221 arabidopsis
32	1298	21.0	1109	10	P93194	P93194 arabidopsis
33	1298	21.0	1110	10	O9F177	O9F177 arabidopsis
34	1293.5	20.9	1098	10	O65440	O65440 arabidopsis
35	1291.5	20.9	1098	10	O8MSK7	O8MSK7 arabidopsis
36	1281.5	20.7	1178	10	O942T3	O942T3 oryza sativ
37	1279.5	20.7	981	10	O9MCA8	O9MCA8 glycine max
38	1279	20.7	994	10	O8SB35	O8SB35 oryza sativ
39	1275.5	20.6	1130	10	O8SB68	O8SB68 oryza sativ
40	1273	20.6	1123	10	O9S819	O9S819 arabidopsis
41	1262	20.4	1041	10	O9F115	O9F115 arabidopsis
42	1260.5	20.4	1145	10	O9FPU2	O9FPU2 pinus sylve
43	1257.5	20.3	987	10	O9SGP2	O9SGP2 arabidopsis
44	1256.5	20.3	987	10	O9MCA7	O9MCA7 glycine max
45	1255	20.3	1002	10	O9AK07	O9AK07 oryza sativ

## ALIGNMENTS

RESULT 1

O22476 PRELIMINARY; PRT; 1196 AA.

AC O22476; 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Brassinosteroid insensitive 1 (Brassinosteroid insensitive 1 gene) (BR11).

GN BR11 OR F23K16.30 OR AT4G39400.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Magnoliophyta; Streptophyta; Embryophyta; Tracheophyta; Eudicotyledones; core eudicots; Rosidae; Spermatophyta; Brassicales; Brassicaceae; Arabidopsids.

OC Eukaryota; Viridiplantae; Magnoliophyta; Streptophyta; Embryophyta; Tracheophyta; Eudicotyledones; core eudicots; Rosidae; Spermatophyta; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

[1]

SEQUENCE FROM N.A.

RC STRAIN=CV; COL-O;

RK MEDLINE=97442355; PubMed=9298904;

RA Li J., Chory J.,

RT "A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction."

RT Cell 90:929-938 (1997).

[2]

SEQUENCE FROM N.A.

RA Beyer M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Scheller C.,

RU Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RU Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

[4]

SEQUENCE FROM N.A.

RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X., Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5]

SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AF017056; AAC49810.1; -  
 DR EMBL; AL078620; CAB4675.1; -  
 DR EMBL; AL161595; CAB80603.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; IRR.  
 DR InterPro; IPR003592; IRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR Pfam; PF00560; IRR; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PRO0019; LEURICHRPT.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; IRR; 15.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.  
 Query Match 99.3%; Score 6140; DB 10; Length 1196;  
 Best Local Similarity 99.4%; Pred. No. 0;  
 Matches 1189; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 MKTSSFFLSTVTLFFSFPSLSFQASPSQSLYREIHQSLSFQVLPDKNLLPDMSSNNX 60  
 DB 1 MKTSSFFLSTVTLFFSFPSLSFQASPSQSLYREIHQSLSFQVLPDKNLLPDMSSNNX 60  
 QY 61 PCTFDGVTCTCRDVKYTSIDLSKPLNVGFSAVSSLSLTGLESFLFNSHINNSVSGFK 120  
 DB 61 PCTFDGVTCTCRDVKYTSIDLSKPLNVGFSAVSSLSLTGLESFLFNSHINNSVSGFK 120  
 QY 121 SASLTSIDLSNSLSGPTVTLTSLGSCGKLFNVSSNTLDPFKVSGGLKNSLEVL 180  
 DB 121 SASLTSIDLSNSLSGPTVTLTSLGSCGKLFNVSSNTLDPFKVSGGLKNSLEVL 180  
 QY 181 SANSISGANVGVWVSDCGELKHLAISGNKISGDVIVRCVNLKPLDVSSNNSTGIFP 240  
 DB 181 SANSISGANVGVWVSDCGELKHLAISGNKISGDVIVRCVNLKPLDVSSNNSTGIFP 240  
 QY 241 LGDGCALDHLDSGNKLSGDFSRRASTCTELKLNISNOFVGPPLPLKSLLYLSLAE 300  
 DB 241 LGDGCALDHLDSGNKLSGDFSRRASTCTELKLNISNOFVGPPLPLKSLLYLSLAE 300  
 QY 301 NKFTGEIPDFLSGACDTLTGLDLSGNHFYGANVPPFGSCILLESIALSSNNFSGELPMDT 360  
 DB 301 NKFTGEIPDFLSGACDTLTGLDLSGNHFYGANVPPFGSCILLESIALSSNNFSGELPMDT 360  
 QY 361 LKMGGLKVLDSFNFSGELPESITLNSASLLTDLSSNNFSGITLNLQNPNTLQOE 420  
 DB 361 LKMGGLKVLDSFNFSGELPESITLNSASLLTDLSSNNFSGITLNLQNPNTLQOE 420  
 QY 421 LYLQNNGFPGKLPPLTNSGSELVSLHSFNYLSGTPISLSGLSKLRDLKLTMLNLEGEI 480  
 DB 421 LYLQNNGFPGKLPPLTNSGSELVSLHSFNYLSGTPISLSGLSKLRDLKLTMLNLEGEI 480  
 QY 481 POELMVVKTLETLIDFNDLTGEIPSGLSNCTNLNWLISLNNRLTGEIPIKWIENLAI 540  
 DB 481 POELMVVKTLETLIDFNDLTGEIPSGLSNCTNLNWLISLNNRLTGEIPIKWIENLAI 540  
 QY 541 LKLSNNSFSGNIPDELGDGSLIWLIDNLNLRGTIPAMFQSGGILANFTAGRYVYI 600  
 DB 541 LKLSNNSFSGNIPDELGDGSLIWLIDNLNLRGTIPAMFQSGGILANFTAGRYVYI 600  
 QY 601 KNDGKKECHGAGNLEFQGISREQLNRLSTRPNCITGRVYGHTSPFDNNGSNMPLD 660  
 DB 601 KNDGKKECHGAGNLEFQGISREQLNRLSTRPNCITGRVYGHTSPFDNNGSNMPLD 660  
 QY 661 MSYNMLSGVLPKEIGSNPFLILNLGHNDISGSI PDEVGDLRGLNLLDLSNKLGRIRIQ 720  
 DB 661 MSYNMLSGVLPKEIGSNPFLILNLGHNDISGSI PDEVGDLRGLNLLDLSNKLGRIRIQ 720

QY 721 AMGALTTLTIDLSNNNLSGPIPEMGOFETPPPAKFLNPGLCGYLPKCDPSNADGYAH 780  
 DB 721 AMSALTTLTIDLSNNNLSGPIPEMGOFETPPPAKFLNPGLCGYLPKCDPSNADGYAH 780  
 QY 781 HOSHGRPPASLAGSVAMGLLFSFVCTFGLILVGRBMRKRRRKEALEMYAEGHNSGD 840  
 DB 781 HOSHGRPPASLAGSVAMGLLFSFVCTFGLILVGRBMRKRRRKEALEMYAEGHNSGD 840  
 QY 841 RTANNTWKLTGYKEALISINLAAPFKPRLTLPADLLQATNGHNSLIGSGGFGDYKA 900  
 DB 841 RTANNTWKLTGYKEALISINLAAPFKPRLTLPADLLQATNGHNSLIGSGGFGDYKA 900  
 QY 901 LKDGSAVALIKLHVSGGQDREPMEMETIGKI KHRNLVPLLYGCKVGBERLLVNEVMK 960  
 DB 901 LKDGSAVALIKLHVSGGQDREPMEMETIGKI KHRNLVPLLYGCKVGBERLLVNEVMK 960  
 QY 961 YGSLLEDVLPDPKKGKVKLKLSTRKKAIGSARGLAFHNCSPHIIHRDKSSNVLIDEN 1020  
 DB 961 YGSLLEDVLPDPKKGKVKLKLSTRKKAIGSARGLAFHNCSPHIIHRDKSSNVLIDEN 1020  
 QY 1021 LEARVSDFGMARLMSANDTHLSVSTAGTGPVPEYVYOSFRCSTKGDVYSYGVLLLEL 1080  
 DB 1021 LEARVSDFGMARLMSANDTHLSVSTAGTGPVPEYVYOSFRCSTKGDVYSYGVLLLEL 1080  
 QY 1081 TGRKPTDSPDGDNNLVGVWQKAKLRISDFPELMEKEDPALEILLQHLKVAACLDD 1140  
 DB 1081 TGRKPTDSPDGDNNLVGVWQKAKLRISDFPELMEKEDPALEILLQHLKVAACLDD 1140  
 QY 1141 RAMRRPTVQVWAFKEIQAGSGIDSGSTIRSIDGGSTLEWDMSEIKVEPEKL 1196  
 DB 1141 RAMRRPTVQVWAFKEIQAGSGIDSGSTIRSIDGGSTLEWDMSEIKVEPEKL 1196

RESULT 2  
 ID 0942F3 PRELIMINARY; PRT; 1121 AA.  
 AC 0942F3;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Putative brassinosteroid-insensitive protein Br11.  
 GN P0480C01.13.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriocarpaceae; Oryzaceae; Oryza.  
 OC NCBI\_TaxId=4530;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone: P0480C01.";  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AP003453; BAB68053.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; IRR.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00560; IRR; 19.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; UNKNOWN\_1.  
 DR ATP-binding; Transferase.  
 QY SEQUENCE 1121 AA; 120180 MM; F71A49845B0E2D09 CRC64;  
 Query Match 49.3%; Score 3050.5; DB 10; Length 1121;  
 Best Local Similarity 53.9%; Pred. No. 1.2e-190;  
 Matches 632; Conservative 166; Mismatches 276; Indels 99; Gaps 20;

Dh	31	LEEFROAVENQAAALXGMSGCGDACHFPAGCGCNGRLTSLSLAGPVLNMFPAVAATTLQ	90
Qy	99	TGLESLEFLSNHSINSVS---GFKCSASITSLDLSRN-SLSGPVTTTSLIGS-CSGKFL	153
Dh	91	GSVEVLSLFGANVSGALSAAGARGCSKLOALDLSGNALRGSVADVAALASACGKXTL	150
Qy	154	NYSNSWTLDPGGKY--SGGKJNSLEVLNLSNLSISGANVGVWYLSDCGELKHALISGNK	211
Dh	151	NLSGAVG-AAKVGCGGGGPGFAGLDSLDSNNKJTTDSDLEMMVADAGAVRWMLDLNR	209
Qy	212	ISGDVVRCSVNLLEFLDVSSNNFSTGIPFLGCSALQHLDISGNKLSGDF-SRAISTCTE	270
Dh	210	IS-----GVEFNNCSGLQYLDLSGMLYGEVPGGALSPCRG	246
Qy	271	LKTLNLSNQFVGPVLPPLPLKSLQYLSLAENKFTGEIPDLFSGACDTLTGHLDSGNHFG	330
Dh	247	LKVLN-----LSFNHLAG	259
Qy	331	AVPPEFGCSLLESIALSSNNPSELPMDTLLMKRGKVLDDLSTNESGELPSELNLSA	390
Dh	260	VFPPIAGLTSLMNLNLSNNPSELPEEAPAKIQOYLTALSLSTFNHFGSLPDTVASL-P	318
Qy	391	SLLTLLDLSNNPFGPILENLCONPNTLOEYLONNGPTGKIPPTLNCNCEVLSHLSFN	450
Dh	319	ELQOQDLSNNFSGTIPBSLQDPNSKHLHYLQNNYLTGIPPAVANCJSLVSLDLSLN	378
Qy	451	YLSGTIPBSLCSLSKRLDKLMLNMLEGEIPQELMYVKTLETLILDNDYLTGEIPSGLSN	510
Dh	379	YINGSIIPASLDGLNLODILIMQNELBEGEIPASLSRIQGEHLILDYNGLTGSIPELAK	438
Qy	511	CTNLNWLSLSNRLTGEIPKMGIGLEMLNLIYKLSNNSPFGNIPELDCRSGLIWDINTN	570
Dh	439	CTKLNWLISLASNRLSGPIPSWLGKSLYALIKLSNNSPFGPIPELDCOSIWLIDLNSN	498
Qy	571	LFNGTIPAMKQSGKLAANFIAPKRAYIYNDDMKKECGAGNLEFQGIREEQNLRLS	630
Dh	499	QNLGSIPELAKQSGKKNVGLIYVRPYVLYRNDLSECEGKQSLLEFSTIRPDDLSRMP	558
Qy	631	TRNPNITSRVYGGHTSPTEDDNNSMMFLDMSVYMLSGYIPKEIGSMPLYFIINTLGHNDI	690
Dh	559	SKKLCNFT-RMYVGSTEXTFKNGSMWFLDLSYQNDLSAIPGELGDMFYIMINLGHNLL	617
Qy	691	SGSIPDEVGDLRGNIIDLSNKLDGFIPOAMGALNLTETIDLSNNNLSGPIDEMGQFET	750
Dh	618	SGTIPSLAEKKLAVIDLSYQNEGPIPNPSFALS-LSBINLSNNQNLNTGIPELGSLAT	676
Qy	751	FPPAKFLNPNELGCPYLPDPCDPSAADQYAHHQRSHGRPRASLAGSVAMGLSFPVCIFGL	810
Dh	677	FPKQYENNTGLCFPLPDCDHSPPRSSNDHQ-SH-RRQSMASSTIAMGLSFLFCIYI	734
Qy	811	ILVGREMRKRRKKKEAL--EMVAEHGNSGDRPTANTNKR--LTGVKEALSLNLAPEK	866
Dh	735	IIAIGSKRRRLKRNKEASTSDIYDSHS--ATNMSDMRKWLSGT-NILSLNLAPEK	790
Qy	867	PLKLTETADLLQATNGHNSLIGSGGFEDVYAAIILKDSAVAIIKLIHVSQGGREFFNA	926
Dh	791	PLQNLTLADIVEATNGFHIAQIGSGGFQDVYAKQKDKGVAAIKKLIHVSQGGREFFNA	850
Qy	927	EMETIGIKIRRNVLPLLGYCKVDDERLNVNWKYGSLEBYLODPKKGVKLKLSTRKCI	986
Dh	851	EMETIGIKIRRNLPPLGYCKABEERLYVDYKFFSLEBYLADRKKIIRKKNWEARRKI	910
Qy	987	AIGSARGLAFLHNHCSPHIITHRMKSSNNVLLDNLEBARVSPFGMALMSAMDTHLSVSTL	1044
Dh	911	AVGARGELATLHNHCIPHIIRHMKSSNNVLLDQULARVSDPFGMARLMSVVDTHLSVSTL	970
Qy	1047	AGTPTGYIPPEYIOSFRCSITGADVSYGVVLLLELLTGKRPDTSDFG-DNNLVGWYKQAK	1105
Dh	971	AGTPTGYVPEPYIOSFRCTTGGDVSYGVVLLLELLTGKRPDTSADFEDNNLVGWYKQHTK	1033
Qy	1106	LRTSDVDPELMKEDPALTELLQHLKVAAACLDDBAKRRPTMVQYMAFKSTIQAGSGID	1165

```
Db      1031 LKIDVDPPELLKEDPVELLELLEHLKINACACLDPRSPRPTMLKVMMFKEIQAGSTVD 1090
Qy      1166 SQSf---IRSIDGGRSTIEMVMDSIKVEYDEEK 1195
          :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1091 SKTSSAAGSIDEGGYG--VLDMPIREAKEER 1120

RESULT 3
Q9ARC8 PRELIMINARY; PRT; 1192 AA.
ID      Q9ARC8
AC      Q9ARC8;
DT      01-JUN-2001 (TrEMBLrel. 17, Created)
DT      01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Hypothetical 129.9 kDa protein.
OS      Lycopersicon esculentum (Tomato).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
CK      Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
NX      NCBI_TaxId=4081;
[1]
SEQUENCE FROM N.A.
MEDLINE=21178822; PubMed=11283350.
RX      Roseberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,
RA      Schumacher K., Schmitz G., Schmid R.;
RT      "Comparative sequence analysis reveals extensive microcolinearity in
RL      the lateral suppressor regions of the tomato, Arabidopsis and Capsella
RM      genomes.";
RC      Plant Cell 13:979-988 (2001).
CC      1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR      EMBL; AJ303345; CAC36401.1; -.
DR      InterPro; IPR000719; Btk_pkinase.
DR      InterPro; IPR001611; LRK.
DR      InterPro; IPR002290; Ser_thr_Pkinase.
DR      InterPro; IPR001245; Tyr_Pkinase.
DR      Pfam; PF00560; LRK; 21.
DR      Pfam; PF00069; pkinase; 1.
DR      PRINTS; PR00019; LEURICHRPT.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PDO00001; Btk_pkinase; 1.
DR      SMART; SMO0220; S_TKC; 1.
DR      SMART; SMO0219; TYRKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW      ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
KM      Transfeaae.
SQ      SEQUENCE   1192 AA;  129941 MW;  BDE1CEDAF8930866 CRC64;

Query Match              42.8%; Score 2643.5; DB 10; Length 1192;
Best Local Similarity    49.5%; Pred. No. 5.2e-164;
Matches 573; Conservative 193; Mismatches 341; Indels 51; Gaps 27;

Qy      35 EIHQQLISFFK-DVLPDKN-LIPDWS-SNKNPCTPDVTGCRDDKYTSIDLSRPKLVAFSA 90
          :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      45 QVGRIILAFAFKKSVEESDPGFLENWTLSSSSPCTMGNISCGNGVVELNLSS-----VGLSG 100

Qy      91 V--SSSLSLTLGEELFLFSNSHINGSVGGFKCASLTSLDLSENLSIGPVTTILTISGSCS 148
          ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      101 LHLTLDMALPTLLRVNFGNHFFGNLSIASCSPEFDLDLSANNFSEVLVLEPLLKSCD 160

Qy      149 GKFLFNVSNTLTDPFGKXVSGGLKLN-SLEVLYDISANSISGANVGVAVLSDGCCELKHAI 207
          :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      161 NKITLVNANSNI-KGVV--LKGPSLIQLDDISSNTISDFGLSTALSN-CQNLTLMFP 214

Qy      208 SGNKISGDV-DVSKCVNLEFLDVSSNNPFSTGP--FLDGCSALQHLDISGNKL-SGDPS 262
          :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      215 SSNKTIAGKKAKSSISSCSLSVLDLSRRNLTGELANDDLGTCONLYTLNLSFNNLTVSVERP 274

Qy      263 RAISTCTEKLINTISSNQFVPPIP--LPLKSLQYLISLAENKFTEGIPDLPGACDTLT 319
          :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      275 PSIIANCOSINTLTIAHNSIRMEIVFEVLVTKLTKSLKRLVLAHNQFDFKIPSELGGSCSTLE 334
```

QY 320 GLDLSGNHFGYAVPFFGSCSLLESALSSNNPSEGLPMDTLLMKRGLKYLDSFNEPFG 379  
 DB 335 ELDSLGRNLTGELPSTFKLCSLFLNLGNNEELSGDFLNTVLSLTLYLYPFPNNITG 394  
 QY 380 ELPLSLNLSASLTLLDLSNNPFGPILPNLCQPKN-TLOELYLQNNGTGKIPPLSN 438  
 DB 395 YPKSLVNCV-KLOVLDLSNNAFIGNVPSFECFPAASGFPLEMTLLASNYLTGVPKQLG 453  
 QY 439 CSELYSLHLSFNYSLSGTIPSSLSGLSKLRDLKMLNMLEGEPQEL-MVYKTLLETLIDF 497  
 DB 454 CRNLRKIDSLNNLVGSLPLETWLPMNLSELVWANNLTGEPGICINGANLQTLTIAN 513  
 QY 498 NDLTGEIPGICNCTNANMISLNNRLTGERPKWIGLENTALIKLSNNSPFGNIPDELG 557  
 DB 514 NFISGTLFQSIKCTNLVWVSLNSRSLGELPQIGINLANIALIQLNNSLTGEPGLG 573  
 QY 558 DCRSLIWLDTNLNFGNITPAMAFKQSGKIANFAGKRYVYIKNDCKECCAGNMLE 617  
 DB 574 SCRNLTIWLDLSNALTGSLPLELDQAGHVPAGASGKQAFVYNEG-GTECRGAGGLVE 632  
 QY 618 FQGRISQNLRLSTPNPCNITSRYVGGHTSPFPDNNSSMFLDMSYNNLSGYIPKEIGSM 677  
 DB 633 FEGIRREKRLILPMVHFCPST-RIVSGRTWYTFISNGSMIYLDLSYNSLSGTIPDNLSL 691  
 QY 678 PYLFTLANCHNDISGSIPEDEVDLRLGMLIILDSNKLGRIPQMSALNLTMLTELDLNN 737  
 DB 692 SFLQVLANGHNNFTGTIPFNGFGKIVGLVLDLSHNSLQGFIPSLGGLSFLSDLDVSN 751  
 QY 738 LSGPPEMGOFETFPFAKFLNPGLCGYPLPCRDPSNADGYAHHQRS--HGRPASLAG 794  
 DB 752 LSGTIPSGQLTTPASRYENNSGLCYPLPPC--GSGNGHSSSLYHNGNKKPTTIG 807  
 QY 795 SVAMGLPSPVCFGLILVGRMRRRKKEALEMYAEHGENSGDRPAANTNKLGYK 854  
 DB 808 NV-VGIMVSFICITILVIALYKIKTONEE-KRDKYIDSLPISG--SSWMLSTVP 860  
 QY 855 EALSTINLAFFERPLKLPADLOATNGPHNDLSGSGFGDYVYAILKQSAVAIKLI 914  
 DB 861 EPLSLNVAFFERPLKLPADLOATNGPHNDLSGSGFGDYVYAILKQSAVAIKLI 920  
 QY 915 HVSQGDREFMAEMETIGIKIRNLVPLLYGCKVDDELLVNEVWKYSLDVLDPKKG 974  
 DB 921 HVTGQDRFMAEMETIGIKIRNLVPLLYGCKVDDELLVNEVWKYSLDVLDPKKG 980  
 QY 975 GYKXKLSRKRKIALGARGLAFLHNCSPHIIHRDKSSNVLLDENLEARYSDPFGMALM 1034  
 DB 981 GMLDLPARKKIALGARGLAFLHNCSPHIIHRDKSSNVLLDENLEARYSDPFGMALM 1040  
 QY 1035 SAMDTHLSVSTLAGTPGYVPEYVYOSFRCTGADVYSYGVVLELITGKRPDPSDFG-D 1093  
 DB 1041 NALDTHLSVSTLAGTPGYVPEYVYOSFRCTGADVYSYGVVLELITGKRPDPSDFG-D 1100  
 QY 1094 NNLVGVWKO-HAKLRISDVDFDELMEKDPALTEILLQHLKVAACIDRAMRPTMVQYM 1152  
 DB 1101 NNLVGVWKO-HAKLRISDVDFDELMEKDPALTEILLQHLKVAACIDRAMRPTMVQYM 1159  
 QY 1153 AMFKETQASGSDISGTI 1170  
 DB 1160 TKEKEVOT---DSESDI 1173  
 RESULT 4  
 Q9LJF3 PRELIMINARY; PRT; 1164 AA.  
 AC Q9LJF3; 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-UTN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Receptor protein kinase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosidie II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=20363099; PubMed=10907853;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,  
 RT TAC and BAC clones."  
 RT DNA Res. 7:217-221 (2000).  
 CC -1. SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AP000603; BAB01743.1;  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR out.  
 DR InterPro; IPR002290; Ser thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00560; LRR; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 14.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 1164 AA; 12660 MW; 79380581D400EEBC CRC64;  
 Query Match 42.6%; Score 2633.5; DB 10; Length 1164;  
 Best local similarity 48.8%; Pred. No. 2,3e-163;  
 Matches 568; Conservative 182; Mismatches 296; Indels 119; Gaps 27;  
 QY 50 NLLPDM--SNKNCPCFDEVTGRD-KYTSIDLSSKPLNVPASAV--SSSLSLTGLSELT 104  
 DB 51 NPLGNRYGSGRDPCTWRVYSCSDGRVIGLDLR---NGSLTGLTINLNLALNSKSL 106  
 QY 105 FLNSHINSVSGFCSASLTSLDSRNSLSGPVITLNLGSCSGKPLNVSNTLDPFG 164  
 DB 107 YLQGNNS--SDSSSSS-----GC----- 124  
 QY 165 KVSGLKLSLEVLDLSANSISGANVYGVLDGCGELKHLAISGNKISGDDVDSRCVN- 223  
 DB 125 -----SLEVLDLSNSNLSIDSVYVFSR-CINLVSVNFSHKLAGKLSSPSASN 174  
 QY 224 --LEFLDVSSNNPSTGP--FLGDC-SALOHLDSGNKLSGFSR----- 263  
 DB 175 KRITTVLDSNNRRSDLEPFTIADFPNSLKHLDLSGNNTGDPFSRLSPGLCENLTVFSLS 234  
 QY 264 -----AISTCTELKLANISSNOFVPIPLP-----LKSLOYLSLAKNKFTGEIP 308  
 DB 235 QNLSGDRFPVSLSNCKLLETMLNLSRNSLIGKIPGDYGNFONLROLSLAHNLVSGEIP 294  
 QY 309 DFLSGACDTLGLDLSGNHFGYAVPFFGSCSLLESALSSNNPSEGLPMDTLLMKRGLX 368  
 DB 295 PELSLCRTLEVLDDLSGNSLTGQLPQSFYSCGLQSLNNGNNLSGDFLSTVVSKLRLIT 354  
 QY 369 VLDSNPFSGELPESLTNLSASLTLLDLSNNPFGPILPNLCQ-NPKNTLOELYQNNG 427  
 DB 355 NLYLPFNNGISVPSILNCS-NKRVLDLSNNEFTGEVPSGFSLOSSSVLEKLLANNY 413  
 QY 428 FTGKIPTLNSCELSVLSHLSFNYSLSGTIPSSLSGLSKLRDLKMLNMLEGEPQEL-MY 486  
 DB 414 LSGVPEVPELGGCKSLKLTIDSLFNALTGLIPKEIWTLPKLSDLVWANNLTGIGIPESICVD 473  
 QY 487 VKTLETLIDFNDLTGELPSGLSCTNLMNLSNNRNLGEIRKWIQRLENLAILKLSNN 546



```

Db 474 GANETLILNNLNLGTSIPESISCKTNMLWISLSSNLLTGIPVIGIKETLALITLQNN 533
Qy 547 SPSSNIDELDCRSLIWLIDLNTNLFNGLTIPAMFKOSGKIAPAFIAGKRYVYKJNDGM 606
Db 534 SLTGNIPSELGNCNLLIWLIDLNSNLLNLGPNGLASQALVPSVSGKOPAFRNBG-G 552
Qy 607 KECHGAGNLEFQCIQISEQNLRLSTNPNCNTSVYSGHTSPFTDNGSNMFLDMSYNML 666
Db 593 TDCGAGGLVEFEGIRARERLEHFPVMSHSCPKT-RIYSGMTVMYMSNGSMIYDLSTNAV 651
Qy 667 SGVPEKIGSMPLFIINTLGHNDISGIPDEVDLRLNLIDLSNLTDRIGQMSALT 726
Db 652 SGSTPLGAGMGYQVANTLGNLITGTPSPFGGLKAGVLDLSDHNDLQGLPGLSGLS 711
Qy 727 MLTEIDLSNNNLSPGIPDEMOQFETPPAKTLNPGLCGYPLPCDPSNADYAHQSHG 786
Db 712 FLSDLDVSNNNLTPIPFGGLTFTPLTRYANNNGLCVPLPCSSG-----SRPRSHA 766
Qy 787 R-RPASAAGVAMGLLSPFCITLGLIIVGRBMRKRKKEALEMTAEGHNSGDRTPANN 845
Db 767 HPKQSTATATMSAGIVSFMCITWLIWALYARK-VQKKEKREKTESLPTSG-----S 820
Qy 846 TNNKLTGVKALSLINLAFAEKPLRKLTFADLTQATNGFNHDSLISGGREGDYKALIKG 905
Db 821 SSMKLSVHEPLSINAVATFKPLRKLTFALHLEATNGFSADSMISGGFGDYKALADG 880
Qy 906 SAVAIKKLHVSGQGRREFMAEMETIGIKHRLNVLPLGYCKVDERLLVNEVMKYSLE 965
Db 881 SVAIKKLIVQCGDRREFMAEMETIGIKHRLNVLPLGYCKVDERLLVNEVMKYSLE 940
Qy 966 DVLQD-PRKSGVTLKLTSTRKINIGSRGLAFIHNHNSPHIIRDMKSNVLTIDENLEAR 1024
Db 941 TVLHEKTKGGITLDMASARKKIALGARGLAFIHSCTPIIHRDMKSNVLTIDOPFVAR 1000
Qy 1025 VSDPGMARLMSANDTLSTLGTAGTGYVPEYXQSFRCSTKGDVSYGVVLELITLGR 1084
Db 1001 VSDPGMARLMSANDTLSTLGTAGTGYVPEYXQSFRCSTKGDVSYGVVLELITLGR 1060
Qy 1085 PMSDPFG-DNNIVGVWKO-HAKLRISDVDPPELMKEDPALIEILLKVAACLDLRA 1142
Db 1061 PTDPEFGEEDNNLVGAKOLYREKGAETILDPETLV-TDRSGVLELHYLKINSQCLDRP 1119
Qy 1143 WRPTVGVWAMPEKE-IOAGSGIDS 1166
Db 1120 FKRPMTIQVMTMFKELVQVDTENDS 1144

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Shin P.P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,
Walker M.W., Altafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,
Gonzalez A.A., Hansen N.N.F., Huizar L.L., Kremetskaya I.I.,
Lenz C.C., Li J.J., Liu S.S., Lucos S.S., Rowley D.D., Schwartz J.J.,
Toriumi M.M., Vysotskaya V.V., Yu G.G., Davis R.R.W.,
Federspiel N.N.A., Theologis A.A., Ecker J.J.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
Khan S., Kim C., Altafi H., Bel B., Chin C., Chou J., Choi E., Conn L.,
Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharbeky N.,
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
Theologis A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC002328; AAF79510.1; -.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR004040; Tyr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00560; LRR; 21.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00370; LRR; 16.
DR SMART; SM00221; STYKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1166 AA; 127423 MW; 8C4DD9231A466A7 CRC64;

```

Query Match 42.4%; Score 2623.5; DB 10; Length 1166;  
 Best Local Similarity 48.5%; Pred. No. 1e-162;  
 Matches 573; Conservative 190; Mismatches 341; Indels 77; Gaps 29;

```

Qy 33 YREIHQISFK--DVLPD-KNLLPDW--SSNKNPCFEDGVTGCD-D-KVMSIDLSKPLNV 86
Db 32 FNEFALLIAKQNSVSDPNVNVGNMKYSGRSGCSWRGVSGDDRIYGLDIRNSGLTG 91
Qy 87 GFSAVSSSLISTGLSPLNSHIN--GVSQGFKCSASITSLDLSRNSLSGPTVTLTSL 144
Db 92 TLNLV--NLTLALPNLQNLVYQGNVFPSSGDDSGSDC--YLQVLDLSSNSISDYSNWDYF 147
Qy 145 GSGSGKLFNVSNNTLDFGKYV--GGYKANSLEVDLSANISGANVGVWVLSGCGEIK 203
Db 148 SKSNLVSIVSNKKV--VKGGFAPSSQSLTYVLDLSTYNIISD-KIPRSFISDPASIK 204
Qy 204 HLAISNKISGVDVRCVNLLEFLDVSNNNFSTGIPFLDGCALQHLIDISGKLSGD-PS 262
Db 205 YLDLTHNNLSGD-----FSDLS-----FQICNLTFFPSLSONNLSGDFRP 244
Qy 263 RAISTCTELKLNISNQVGPFP--LPLKSLQVLSLAKNKFTGETIPDLGACDTL 318

```

Db 245 ITLPNCKELETINISRNRLAKGIPNGRWGSPFNKQKLSLHNRISGEIPELISLCKTL 304  
 QY 319 TGLDLSGNHFGVAVPPFGSCSLLESALSSNNFSGELPMDTLTKRGKLVLDLSENEFS 378  
 Db 305 VIIDLSSNTEGSELPISQFACVWLQNLNNTLNNLSGDFLTVASKITGITYLVAVANNIS 364  
 QY 379 GELPESTLNNLSASLTLTDLSSNNFSGELPMDTLTKRGKLVLDLSENEFS 437  
 Db 365 GSIVISLTLNCS-NIRVYDLSSNGTGVNPGFSGLOSSPVLEKLIANNVLSGVPELQ 423  
 QY 438 NCSELYSLHSFNYLSGTPSSLSLSKRDKLMNMLSEIPELMYK--TLETLLI 495  
 Db 424 KCKSLKTLIDLSFNLGELPISQFACVWLQNLNNTLNNLSGDFLTVASKITGITYLVAVANNIS 482  
 QY 496 DPNLDTGEIPEGLSNCNTLNNLSISNNRNLGELPISQFACVWLQNLNNTLNNLSGDFLTVASKITGITYLVAVANNIS 555  
 Db 483 NNNLTLGSPISIRCTNMWISLSSNNLTKIPSGIGLISKLILOLGNNSLSGNVPRQ 542  
 QY 556 LGDCRSILWLDLNTNLFNGITIPAMFQSGKIAANFLAGKRYVYIKNDGMRKECHGAGNL 615  
 Db 543 LQCKSLTLWLDLNNLNLGDLPEGLASQAGLVMPGVSQGFAYVRNEG-GTDCRAGGL 601  
 QY 616 LEFGIRSEQLNRLSTNPNCNITSRYVIGHTSPFPDNNSSMFLDMSNMISGTYIPEIG 675  
 Db 602 VEFEGIRAEERLELPMVHSCPAT-RIVSGMTYTFPSANGSMIYEDISYNAVSGFIPGCG 660  
 QY 676 SMPYFIINLGNNDISGIPDEVDGLNLTLNLSNKLDRIPQAMSLTMLTEIDLSN 735  
 Db 661 NNGYLVQVNLGNRLTGITIPDSFGELKALIVLDLSHNVLQGYLGSLSGLSFLDLVDN 720  
 QY 736 NNLISGPIDEMGQFETPPAKFLNNPGLCGYPLPRCDPSNADGVAHQRSGRRPASLAGS 795  
 Db 721 NNLTPIPRPGGLTFFPSRYANNGLCGVPLRPC--GSAPRRPITSRIHAKK-QYVATA 777  
 QY 796 VAMGLIFSVCLFGLILVGRERKRRKKEALEMYAEGHNGSGDRANTNMKLTGVKE 855  
 Db 778 VTAGIAFSFMCVWMLVYVRK-VQKEQKREKITSLSPTSG-----SCSWKLSSVPE 831  
 QY 856 ALSINLAERKFLRKLTFADLLOATNGFNDLSLISGGFGDYVYKAILKDSNAVAIKKLIH 915  
 Db 832 PLISINVAITEKRLKTLFPHLEATNGFSAETMWSGGEGEYVYKQKLDGSAVVAIKKLIH 891  
 QY 916 VSGQGDREFMAEMETIGIKIHNVLVPLIGYGVGBERLIVNEVMKYSLEVDYLD--PKK 973  
 Db 892 ITGGQGRREFMAEMETIGIKIHNVLVPLIGYGVGBERLIVNEVMKYSLEVDYLD--PKK 951  
 QY 974 GGYKLTSLRKLKIAIGSAGLAFLHNCSPHIIHRMKSNNVLDLDENEARVSDGMARL 1033  
 Db 952 GGYIWMARKKIAIGSAGLAFLHNCSPHIIHRMKSNNVLDLDENEARVSDGMARL 1011  
 QY 1034 MSAMDTLHSLVSTAGTPGYVPEYYSFRCSTKGVYSGVVLLELTGKRTPSDPRG- 1092  
 Db 1012 VSLDHLHSLVSTAGTPGYVPEYYSFRCSTKGVYSGVVLLELTGKRTPSDPRG- 1071  
 QY 1093 DNNLVGVWVQ-HAKRISDVPELMEKEDPALEIQLKVAVACTLDRAARRPRMVOY 1151  
 Db 1072 DNNLVGVWVQ-HAKRISDVPELMEKEDPALEIQLKVAVACTLDRAARRPRMVOY 1130  
 QY 1152 MAMFKEIONGSGIDSGSTIRSIDGSPSTEMVMSIKEVP 1192  
 Db 1131 MAMFKEIONGSGIDSGSTIRSIDGSPSTEMVMSIKEVP 1156

## RESULT 6

PRELIMINARY; PRT: 1166 AA.

AC 09ARF3  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DB Hypothetical 127.3 kDa protein.  
 OS Capsella rubella.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Capsella.  
 OK NCBI\_Taxid=81985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21178822; PubMed=11283350;  
 RA Rosenberg M., Theres K., Acarkan A., Herrero R., Schmidt T.,  
 RA Schumacher K., Schmidt G., Schmidt R., extensive microcolinearity in  
 RT "Comparative sequence analysis reveals the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
 RT genomes.";  
 RL Plant Cell 13:979-988(2001).  
 CC -SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL, AJ030349, CACG6390.1; --  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00560; LRR; 18.  
 DR Pfam; PF00063; kinase.1.  
 DR PRINTS; PR00019; LEURICHRP.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;  
 KW transferase.  
 SQ SEQUENCE 1166 AA; 127261 MW; DADE2E2B4C5647A0 CRC64;  
 QY 8 FLSVTLTFFSFFSLSPQASPEQSYREIHQISFK--DVLDP-KQLLPDW--SSNKNPC 62  
 Db 5 WLFVVLICFTLALGHGKRL-INSDEDFALMAFQFQSVKSDPNVVLGNWVYESGRGSC 63  
 QY 63 TPDGVTCRD-KVTSIDLSKPLNFGSAV--SSLSLITGESLPLNSHINGSVSGK 119  
 Db 64 SWRGVSCSDGDRIVGDLR---NGGVTLNLANLTALPNLQNYLYQNYSSSSSGGS 119  
 QY 120 CSAS--LTSIDLSRSLSGPTVTLTSLSCGSLKPLNVSNTLPPGKVS--GGLKLSNL 175  
 Db 120 SSGSYCYLVLTLSNLISLIDYSLVDYVFSKSNLVSNSNKL--VGLGRAPSSLSL 177  
 QY 176 EVLDSANSISGANVVGWVLSDCGELKHLAISGNKISG--DVDVRCVNLLEPLDVSN 232  
 Db 178 TYDPSYNIIS-EKIPESFISEFPASTKYLDTLHNNFSGDFSLSPGMCNLSFFSLSON 236  
 QY 233 NFSTGIPFLDCALQHLDISGNKLSGDPBSRAISTELKLNISNQVGPVPP--L 288  
 Db 237 N-----ISGVK---FPISLNCRFLETLNLSNNLAKGIPGEYWG 274  
 QY 289 PIKSIQVLSAENKFTGEIPDFLPGACDTLTGDLISGNHFGVAVPPFGSCSLLESALS 348  
 Db 275 SFQTLKQLSLHNFSEIPELISLCKITLETIDLSGNLMSGELPQFACVWLQNLNIG 334  
 QY 349 SNNFSGELPMDTLTKMGKLVLDLSENEFSGELPESITNLISASLTLTDLSSNNFSGPILP 408  
 Db 335 NNYLSGDFLSTVSKIRIRIYLYVAFNNISGVISITNCT-NLRVYDLSSNFTGNVPS 393  
 QY 409 NLQ-QNPKNTLQELYLQNGFTGKIPTLSCGEIVSIHSENYLSGTPSSLSKLR 467  
 Db 394 GLCSQSSPVLEKLIANNVLSGTVPELCKSKLTKITIDLSGELPPIKDYVMTLNL 453  
 QY 468 DLKLMNMLSEIPELMYK--TLETLLIDPNLDTGEIPSGLSNCTNLNWLSSNNRLT 525  
 Db 454 DLVMMANNLTGSIPEGVCKGKLETLIIINNNLITISIQSISRCNTNMWISSNRLT 512

QY 526 GEIPKATGRLNTALITLKLNNSSFGNIPDELGDRCSTLIMDLNTLNGTIPAMKFGSG 585  
 Db 513 GKIPGIGNSKLAILQLGNNSSLGNNVPROLGCKSLIMDLNNSNLTGDLPGELASQAG 572  
 QY 586 KIANFIAGKRYVYIKNDGKKKECHAGNLLEPGIRSEDLNLSRNCNTSRVYGGH 645  
 Db 573 LVMGSGVSGQFAVRNEG-CTDRGAGGLVEFEGIAERLEFPVHSCPAT-RVYSGM 630  
 QY 646 TSPFPDNNSSMFLDMSYNNMLSGYIPKEIGSMYPLFLNLGNHDSIGSPIDEVGDRLGN 705  
 Db 631 TMYTFSANGSIWYEDISYNVSGILPGYGMGYLQVNLGNHRTIGNIPDLSGLGKAIG 690  
 QY 706 ILDSSKRLGRILOAMASLMTLTDLSNNNSGPIPEKQGEPIFPKFLNPGLCGY 765  
 Db 691 VLDLSHNDLGGYRPGSGISLFSFLDSDVSNNNLTGPIPGQGLTFPPVRYANNGLCGV 750  
 QY 766 PLPRCDPSNADGYAHQSHGRRP-----ASLAGSANGMLFSFPCIGLLVGRE 816  
 Db 751 PLRPGG-----SAPRRPITSSVHAKQGLATVINGIAPSEFWCLVLMFALYR 798  
 C 817 MKRKRKKKALEMYAEHGNSGDRTPANNNTMKLTGVKALSINTLAEPKPRKLTFAHL 876  
 Db 799 VRK-VCKKELKREKYLESLPTSG-----SCSWKLSVPEPLSINAVTEPKRKLTFAPHL 852  
 QY 877 LOATNGFNDLSLGGSGFGDYVYKALIKDGSNAVAIKLHVSGQGREMAEMETIGKIKH 936  
 Db 853 LEATNGFSAETWYSGSGFGDYVYKALIKDGSNAVAIKLHVSGQGREMAEMETIGKIKH 912  
 QY 937 RNLVPLLYGCKVQDERLVLNVEWVKYSLDVLQD--PKKGVYKLTSTRKALISAGL 994  
 Db 913 RNLVPLLYGCKVQDERLVLNVEWVKYSLDVLQD--PKKGVYKLTSTRKALISAGL 972  
 QY 995 AFLHNHCSPIIHRDKSSNVLDENLEKRVSDFGMARLMSADTLSTVSTLAGTPGYVP 1054  
 Db 973 AFLHNHCSPIIHRDKSSNVLDENLEKRVSDFGMARLMSADTLSTVSTLAGTPGYVP 1032  
 QY 1055 PEYYSQFRCSTKGDVYSGVYVLELITGKRPDPSDFG--DNLVGVWYQ-HAKLRISVF 1112  
 Db 1033 PEYYSQFRCSTKGDVYSGVYVLELITGKRPDPSDFG--DNLVGVWYQ-HAKLRISVF 1092  
 QY 1113 DPELMKEDPALEIELLQHLKVAACLDPRAMRPTMVQVMAFKFIOAGSGIDSGSTIRS 1172  
 Db 1093 DPELMKEDPALEIELLQHLKVAACLDPRAMRPTMVQVMAFKFIOAGSGIDSGSTIRS 1145  
 QY 1173 IBDGGSTIEMDMSIKETP 1192  
 Db 1146 -----SLDEFSLKETP 1156

QY 92PS9 PRELIMINARY; PRT; 1143 AA.  
 AC 092PS9, 01-MAY-1999 (TREMELREL. 10, Created)  
 DT 01-MAY-1999 (TREMELREL. 10, Last sequence update)  
 DE 01-JUN-2002 (TREMELREL. 21, Last annotation update)  
 DR Putative receptor protein kinase.  
 GN AT2G01950.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA,  
 EX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldguy T.V.,  
 RA Buelin C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H., Motilal K.S.,  
 RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eilen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:761-768 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RP  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Bani H., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carrincci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Shinozaki K.,  
 RA Davis R.W., Becker J.R., Theologis A.;  
 RT "Full length cDNA of gene At2g01950 (GI:15226381).";  
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 CC 1-SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AC006532; AAD20088.1; -  
 DR EMBL; AY074313; AAL67010.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; IRR.  
 DR InterPro; IPR003592; IRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR Pfam; PF00560; LRR; 21.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICRPT.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 18.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00106; PROTEIN KINASE ST; 1.  
 KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;  
 KW TRANSFERASE.  
 SQ SEQUENCE 1143 AA; 125675 MW; 7D1C88493F27A94E CRC64;  
 Query Match 40.6%; Score 2509.5; DB 10; Length 1143;  
 Best Local Similarity 46.2%; Pred. No. 2.8e-155;  
 Matches 556; Conservative 183; Mismatches 353; Indels 111; Gaps 25;  
 QY 9 LSATTLFFPFFSLSFQASPSQSLYREIHQISFKCVLPD--RNLPLDWSNNKPPCTPDG 66  
 Db 14 IQISFIFLTLHSGS--SSDSQSLKTLSTLSFKTMIGDDPNNTLSNNSPKSPQFSG 72  
 QY 67 VTCRDDKVTISLSSKPLN--VGFSAVS-----SLSLTGLBSLFLSNHINGSVGFKC 120  
 Db 73 VTCAGRTIELNLSGSGLSIGTGSFNAFTSLDSLVKLS--ENFPLVN-----STSLTL 125  
 QY 121 SASLTSLDLSNLSGVTTLTSLGSCSGKFLNVSNTL---DPPGKYSGGKJLNS--L 175  
 Db 126 PLTLTLELSSGLIG-----TLPENFPSKYNLSLITSLYNNFTGKLPDLFLSSK 179  
 QY 176 EVLDLSNLSIGANVWVWVSDGGLKHLAISGNKISGDVDVSRVNLFLDVSSNFS 235  
 Db 180 QTLDSLNNITIG-----PISGLTIP----- 199  
 QY 236 TGIPFLDGSALQHLDISGNKLSGDFSRALSTETKLTINISNPPVGPPL--PLKSL 293  
 Db 200 -----LSSCVSMYLDPSGNSIGYISDSINCTNLKSLNLSYNNFDGIPSPFELK 254  
 QY 294 QYLSLAENKTKTGIPLPFLGACDTLTGLDLSGNHFYGAVPFPGSCSLSLSLALSSNFS 353  
 Db 255 QSLDSLNRULTGWIPEIGDTCSLQNLRLSYNNFYGIVPESLSSCSMWLQSLDLSNNIS 314  
 QY 354 GELPMDTLKRGKGLDLSFNFGSELPESLTNSA--SLTLTDLSSNNSGPILPMLC 411  
 Db 315 GPFWTLKRSFGLQILLLSNLISGDP--TSISACKSLRIADFSNNRFSGVIPDLIC 371

QY 412 QNPKATLOEYIYLNNGFTGKIPPTLSNCEIYSLHSENYLSGTIPSPSLSKLRLDKL 471  
 Db 372 PGAA-SLEELRRLPDNLVTGIPPAISQCSBLRTIDSLNYLNGTIPPELGNOKLEQFIA 430  
 QY 472 WLMMLKEEIPOLIMYVKTLETLILDENLDTLGEIPSGLSNCTNLMIISLNNRLTGEIPK 531  
 Db 431 WYNNINAGEIPPELIGKQNLKDLILNNQNLGELIPPEFNCNSINEMWFSNRLTGEVPRD 490  
 QY 532 IGLLENLAIILKLSNNSFSNGNIPPELGDRLILMDLNTLNFNTIPAAHFGSG-KIAAN 590  
 Db 491 FGLISRLIAVLQILNNNTGELIPPELGTCTTLVMDLNTLNTLGEIPRIKRGQSFALSG 550  
 QY 591 FIKGKRVVYIKNDGKKECHGAGNLLEFQIRSEQLNRLSTRNPNCTISRYGHTSPTE 650  
 Db 551 LLSGNTMAFPRNVG--NSCGVGVGVFSGIRPERLLQIPSLSCOPT-RMYSGPILSLF 607  
 QY 651 DNNGSMFLDMSYNNLSGYIPKEIGSMPLYFILNLGHNDISGSIPEVGDRLGNTLIDS 710  
 Db 608 TRYQTIETLDLSYQNLGKIPDEIGEMIALQVLELSHOLSGELPFTIGQLKNLGVFDS 667  
 QY 711 SNKLDGRIPOAMSLIMLTELIDSNNNLSGPIPEMGQFETFPFAKFLNNGLGCGYELPC 770  
 Db 668 DNRIQGOIPESFSLSPVQIDLSNNEITOPIPORGOLSTIPATQYANNPGLGVPLPEC 727  
 QY 771 DPSNAD--GYAHHQRS-HGRPRASLAGSVAMGLLFSFVCI FGLIIVGEMKRRKKEA 826  
 Db 728 KNGNNOPLPAGTEBGRKAKHGTBRASWANSIYLVGLISAASVCIILVIAVARARRDAD 787  
 QY 827 ELEMYABGHGNSGDRNTANNWKLTVGEKALSITLAAFEKPIRLCTPADLLOATNGPHND 886  
 Db 788 AKMLHSLOAVNSA-----ITWKIEKEKEPLISINAVFQRLKLFKSQLEATNGSAA 841  
 QY 887 SLISGGGFDVYKAILKQDSAVAIAIKLIHVSQGDREMEMETIGIKIRNLVPLLYGC 946  
 Db 842 SMIGHGGGGEFPAKTIKQSSVAIKKILRLSCQDRBEMAMETLIGIKIRNLVPLLYGC 901  
 QY 947 KYGDERLLVNEWKYSGLEEDVLODPKKGVR-LKLSRKRKIAIGSARGLAFLHNSPH 1004  
 Db 902 KIGERLLVYEMQYSGLEEDVLODPKKGVR-LKLSRKRKIAIGSARGLAFLHNSPH 961  
 QY 1005 ITHRMKSSNVLIDENLARYSDPGMARLMSMDTHLSVSTLAGPGVPEYQSFRC 1064  
 Db 962 ITHRMKSSNVLIDENLARYSDPGMARLMSMDTHLSVSTLAGPGVPEYQSFRC 1021  
 QY 1065 TKGDYVSYGVLLLELTGRPTSDPGDNMLVGVYKQHANL-RISDVPELMKEDPAL 1123  
 Db 1022 AKGDYVSYGVLLLELTGRPTSDPGDNMLVGVYKQHANL-RISDVPELMKEDPAL 1081  
 QY 1124 EI-----ELQHLKVAVALCDDDRAMRRPTVQVWAMKEIOAGSGIDSOSTI 1170  
 Db 1082 SLNEKEGPEGVIVKEMRLYLALRCVDDPFSKRPNNLQVAVASLREUR-GSENNSHSHS 1140  
 QY 1171 RSI 1173  
 Db 1141 NSL 1143

RESULT 8  
 094LN2 PRELIMINARY: PRT: 1110 AA.  
 AC 094LN2;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative receptor protein kinase.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI Taxid=4530;  
 RP [1] SEQUENCE FROM N.A.

RA Wing R.A., Fritsch D., Presting G., Wood T., Yu Y., Soderlund C.,  
 RA Kim H., Rambo T., Henry D., Simmons U.,  
 RT "Rice Genomic Sequence."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC078891; AAK52544.1;  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR InterPro: IPR002290; Ser\_cin\_pkinase.  
 DR Pfam: PF00560; LRR. 19.  
 DR Pfam: PF00069; pkinase. 1.  
 DR ProDom: PD000001; Euk\_pkinase. 1.  
 DR PROSITE: PS00107; PROTEIN KINASE APP, UNKNOWN\_1.  
 DR PROSITE: PS00108; PROTEIN KINASE DOM, 1.  
 DR PROSITE: PS00030; RR\_NF\_1; UNKNOWN\_1.  
 DR ARP-binding; Kinase; Receptor; Transferase.  
 KW SEQUENCE 1110 AA; 118109 MW; B6723380C0A8E94 CRC64;  
 SQ

Query Match 36.2%; Score 2240; DB 10; Length 1110;  
 Best Local Similarity 44.2%; Pred. No. 1, 1e-137;  
 Matches 510; Conservative 172; Mismatches 361; Indels 112; Gaps 30;

QY 39 LISFDVY--PDKNLLPDWSSNNKPTCFDGYTCR-DKVTSIDSKPLAVGSAVSSL 95  
 Db 28 LIRFAFVHKDPRGVLSM--VDPGCRWRTVCNGDRVTELDAA-----GGLAGREL 81  
 QY 96 LSLTLESLFSLNSHINSV---SG--FKCSASLTLSDLSNLSGVTTLTSLGSCGL 150  
 Db 82 AALSGDLTLCRLNSGNEHVLVDGDLVLPALQLDLSGGLAGLPLD----- 131  
 QY 151 KFLANSSNTLDPFGKVSGLKINSLEVLDSANSISGANVGVWVLDGCGELKHLAISGN 210  
 Db 132 GFLACYPNLT-----VSLARNLTG-ELPGMLA--SNIRSFDSGN 171  
 QY 211 KISGVD--VSRVNLFDVSSNNFSTGI-PFLGDCALQHLDISGNLISGDFSRALSTC 268  
 Db 172 NMSGDISGSLPATLAVLDLSGNRFTGAIIPSLSGCAQLTTLNLSYNGLAIPBGICAI 231  
 QY 269 TELKLNLSNNGFVPIPP---LPLKSLQVLSLAENFTGEIPDFLSGACDTLTGLDLS 324  
 Db 232 AGLEVLDVSWNLTHGAIIPGLGRNACALRVLRVSSNISISIPESLS-SCGALRLDVA 290  
 QY 325 GNHFYGAVP-PFPGSCSLSSLSLSSNNPSEGLPMDTLKRGKLVLDLSPREFGELPE 383  
 Db 291 NNVVSGGIPAAVIGNLTAVESLILNNFISGLP-DTIAHCKNLVADLSKISGALP- 348  
 QY 384 SLTNLSASLITLDSNNSGPIIPNLQNPKNLTLOELYLQNGFTGKIPPLSNCELY 443  
 Db 349 -----NELC-SGPAALBELRLPNLVAGTIPPLGSLNCSRLR 383  
 QY 444 SLHLSFNYSLTGPSISLGLSKLRDKLMLNLEBIPOLIMYVKTLETLILDENLDTGE 503  
 Db 384 VIDPSINYLGRIPPELIGRLALREKLVMMFNGLDGRIPADLQGCNRLRTLLINNFIQGD 443  
 QY 504 IPGSLNCTNLNMIISLNNRLTGEIPKMIIGLEBNALILKLSNNSFSNGNIPPELIGRSLI 563  
 Db 444 IPVELFECTGLEWVSLTSNOITGITIRPEGRLSRLAVIQLANNSLAGELPRELGNSSIM 503  
 QY 564 WLDLNTLNFNGTIPAMFQSGKI-AANFLAGKRYVYIKNDGKKECHGAGNLLEFQIR 622  
 Db 504 WLDLNSRLTGEIPRIKRGOLSTPLSGISNTLAFVNVG--NSCKGVGVLEFAGIR 561  
 QY 623 SEQLNRLSTNPNCTISRYVGHSTSPFDNNGSMFLDMSYNNLSGYIRKEIGSMPLYFI 682  
 Db 562 PERLLQVPTIKSCDPT-RIVSGAAGVWTRVOTLEVLIDLSYNSLDEIRPEELADVAVLV 620  
 QY 683 LNLGHNDISGSIPEVGDLAGNLITLDSNKLDRIPOMASALTMLELIDLSNNSLGP 742  
 Db 621 LIDLARNNLITGEIPASIGRLNMLGVFVDSRRRLQGGIPDSNSLSFLVQIDISNNLSGEI 680  
 QY 743 PEMGQFETFPFAKFLNNGLGCGYPLPCDP-----SNADGYAHHQSRGRRPASLAGS 795

DB 681 PORGOLSTLPASQVAGNPGLCMPLEPCGDRIPATMSGLAAASTDPPRAVATWANG 740  
 QY 796 VAMGLIFS--FVCIFGLIIVGEMRRRRKKEALEMYAEGHNSGDRANTNMKL-TG 852  
 DB 741 VILAVIVSAGLCAAIWVAARARREVRSAMLSLQD-----GTRTA--TWKLGKA 793  
 QY 853 VVEALISINLAEEKPLKLTPLADLLQATNGFINDSLISGGGVDYKAILKDGSAVAIKK 912  
 DB 794 EKEALISINVAIRORRLKLTFTOLLFANTNGFSTASLISGGGVEFKATLKDGSAVAIKK 853  
 QY 913 LIIHVSQGGREEMAEEMETIKIKHENIVPLLAGCKYKDEBLLVNEVMKXSLIEDVLQDP- 971  
 DB 854 LIHLVYQGGREEMAEEMETIKIKHENIVPLLAGCKYKDEBLLVNEVMKXSLIEDVLQDP- 913  
 QY 972 -KKGVKCLKLSTRRKIALISAGARLAFLLHNSCPHIIHDMKSSNVLLDENTLBARVDFGM 1030  
 DB 914 GRSASPAWSMEORRKIVAGARGLCFLLHNCIPHIIHDMKSSNVLLDQDMBARVADFGM 973  
 QY 1031 AILMSAMDTLHLSVSTLAGPGVPPREYVOSFRCTKGVYSGVVLLELLTGKRTDSDP 1090  
 DB 974 AILMSAMDTLHLSVSTLAGPGVPPREYVOSFRCTKGVYSGVVLLELLTGKRTDSDP 1033  
 QY 1091 FGDNNIVGVVVKOHAKRLISD-----VFPDLAKEDPALEIELLQIKVAVALCLDRAWR 1145  
 DB 1034 FGDNLVGVV-----KMKYGDGAGKEVLDPELVEGADAD-EMARMDALQCVDDPFSKR 1088  
 QY 1146 PPMVAVAMAEKIQ 1160  
 DB 1089 PNMLOVAMRELDA 1103

RESULT 9  
 Q9LYN8 PRELIMINARY; PRT; 1192 AA.  
 ID Q9LYN8  
 AC Q9LYN8;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Receptor-like protein kinase-like protein.  
 GN T28J14.220.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 CC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 F. Rudd S., Lemcke K., Mayer K.F.X.;  
 F. Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DB EMBL, AL163662, CAB87284.1;  
 DB InterPro; IPR000719; Euk\_kinase.  
 DB InterPro; IPR001611; LRR.  
 DB InterPro; IPR003592; LRR\_out.  
 DB InterPro; IPR002290; Ser\_thr\_kinase.  
 DB InterPro; IPR004040; STY\_kinase.  
 DB InterPro; IPR001245; Tyr\_kinase.  
 DB Pfam; PF00560; LRR; 25.  
 DB Pfam; PF00069; pkinase; 1.  
 DB PRINTS; PRO0019; LRRICHRPT.  
 DB PRINTS; PRO0109; TYRKINASE.  
 DB ProDom; PD000001; Euk\_kinase; 1.  
 DB SMART; SMO0370; LRR; 25.  
 DB SMART; SMO0221; STYK; 1.  
 DB PROSITE; PS00107; PROTEIN KINASE\_ATP; 1.  
 DB PROSITE; PS00011; PROTEIN KINASE\_DOM; 1.  
 DB PROSITE; PS00108; PROTEIN KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 1192 AA; 12979 MW; D02C8B8FE6B09F7D CRC64;  
 Query Match 26.0%; Score 1608.5; DB 10; Length 1192;  
 Best Local Similarity 35.2%; Pred. No. 2.4e-96;  
 Matches 453; Conservative 177; Mismatches 421; Indels 235; Gaps 36;  
 QY 11 VTTLFFSFFSLSPQASPSQLYREIHOLISPKDVLDPDKNLPDW--SSKNKCTFDGYT 68  
 DB 4 LTALFLFLPF--SFSSSAIVDLSEFTSLISFRSLSPNSLLSSNMVSSASHCDVGYT 61  
 QY 69 CRDDKTSIDLSKPLAVNGSAVSSLSLTGLESIFLNSHINSVS---GFKCSASL 124  
 DB 62 CLIGRVNLSLPLSLR--GQIPKEISILKMLRELCLAGNPFSGKIPPEIMWLK---HL 115  
 QY 125 TSLDLSNLSGCVTTLTLSIGSCGKFLNVSNTLDPFGKV--SGGLNLSLEVLDSA 182  
 DB 116 QTDLDGNSLTGLPL--LSLPLQLLYLDLSDN--HFGSLPPEFFISLPALSLDVSN 171  
 QY 183 NSISGANVGVVLSDCGELKHLA--ISGNKISGDV----- 216  
 DB 172 NSLSGE-----IPPEIGKLSNLSNLYMGLNPSGQIPSEIGNISLLKNPAAPSCFNGP 225  
 QY 217 ---DVRGVNLEFLDVSSNNPSTGT-----PFLGDCSALQ 248  
 DB 226 LPKEISKRLAIDLSTYNPLKCSIPKSGELNLSILNLSAELGLIPPEIGCKSLK 285  
 QY 249 HLDISGNKLSGDFRSRAISCTELKLNIS--NOFVGPPL--PLKSLQYLSLENKFT 304  
 DB 286 SLWLSNLSLSPLELS--ELPLTFSAERNQLSGLSPSMGKWKVLDLSLLANNRPS 342  
 QY 305 GEIP-----DFLSGA-----CDTLTGLDLSGNHFGYAVPPFGSCSL 341  
 DB 343 GEIPHEIEDCPMLKHLASLNSLISGSIPELGGSGLEAIDLSGNLSGTIEVVDGCS 402  
 QY 342 LESLASSNNSGELPMDTLKKRGKXVLDLSPNPSGSLPSL--TNL----- 388  
 DB 403 LGELLTNNQINSIPED-LMKL-PLMALDLDSNNFTGRIPLSKWSTNLMFTASYNRL 460  
 QY 389 -----SASILTLSDSSNNPSGPIPLNLCMPKNTLGLVYLNNGFTKIPPTLSN 438  
 DB 461 EGYLPAEIGNAASLKLVLSDNQLTGEIREF--GKLSLSLTANANVFOKRIEVELSD 518  
 QY 439 GSELVSLHSPNYLSTGTPSSLSGLKLDLKLMLNMEGEIPQ-----ELMY 486  
 DB 519 CTSLLTLSDGNNLQGIIPDKITALAQCLVLSYNNLSGSLPSKPSAVFHQIEMPDLSF 578  
 QY 487 VKTLFLLIDPNDLGEIRSGISGNCNTNLMWISLNNRLTGERPKMIGLENTAIIKLSNN 546  
 DB 579 LQHHGIFDLSTYRNLGPIPEELGECIVLSEISLNNHLSGEIPASLSRLTNLTIDLSGN 638  
 QY 547 SFGNIPDELGCPSLIWDLNTLNFNGTIPAMAFKQSGKLAANFIAGRYYYINDQMK 606  
 DB 639 ALTGSLPKEMGSLKLGINTLANNQINGHIP----- 669  
 QY 607 KECHGAGNLLFQGISRQQLNLSTRNPNCTISRYGHTSPTPDNNGSMFELDSYML 666  
 DB 670 -----ESFGLGSLVTKLNTLKNKLDGVPASLGNLKLTLTMDLSFNNL 712  
 QY 667 SGYIPKEISMPYFLTNLGHNDISGTPDEVDRGNTLIDLSNKKLDGRIPOAMSLT 726  
 DB 713 SGELSELSTMEKLVGLYIEQNKFTGEIPSELGNLTQLEVDVSNLISGEIPTKICGLP 772  
 QY 727 MLTEIDLSNNLSGPIPEWGCGETPPPAKFLANPGLCGVPLPCDPSNADGVAHHQSHG 786  
 DB 773 NLEFLNLAKNNLRGEVPSGVQDPSKLLSNKELCG---RVGSGC-----KIEG 821  
 QY 787 RRPASLAGSVAMGLIFS--FVCIFGL--IIVGEMRR--RRKKEALEMYAEG--HG 836  
 DB 822 TKLSAMGIAGMLGFTIIVFVFSLRWMAWKRVKQDDPERNEBSRLGVDQNLVYF 881  
 QY 837 NSGDPTANTNMKLGVYEAALISINLAEEKPLKLTPLADLLQATNGFINDSLISGGGVD 896  
 DB 892 LSGSRS-----RPPISINIAFEQPLKVLKVLGDIIVEATDHFESKNITIDGGGRT 930

QY 897 VYKALDGSVAIAIKLHVSQGDREFFAMMETIGIKERNVLPILIGYKVGDERLLVN 936  
 DB 931 VYKACLPEKTVAAVYKLSBAKTQGNREFFAMMETIGIKERNVLPILIGYKVSFSEKLLVY 930  
 QY 957 EVMKGSJEDVLOPKKGVYKLSKTRKIAISARGIAFLAHNHCSPHIIHRDKSSNVL 1016  
 DB 991 EYMWGSLDHMLRNQTMLEVLWDSKRIKIAVGARGLAFHGHGPTPIIHRDIKASNIL 1050  
 QY 1017 LDENLEAVSDFGMARLMSAMDTHLSVSTLACTPGVPEPYQSFSGSTKGDVYSYGV 1076  
 DB 1051 LDGDEPFPVAFGLARLISACESHVS-TVIACTGYIPEPEYQSGARATTKGDVYSFVIL 1109  
 QY 1077 LELLTKRPTDSDPDGDN--NLVGVYKQAKL-RISVPEPELMKEDPALEILLQHLK 1132  
 DB 1110 LELVYGEKPT--GPDEKSEGGNLVGMALOKINOGKAVDVIDPLV--SVALKNSQLRLQ 1166  
 QY 1133 VAVAGLDDRAMRPTMVOMAMFEKI 1158  
 DB 1167 IAMLCLAEPPAKRPNMLDVALKAKEI 1192

RESULT 10  
 Q8RZV7 PRELIMINARY; PRT; 1294 AA.

AC Q8RZV7  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative receptor-like protein.  
 GN P0413C03.19.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 clone: P0413C03."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF003451; BAB6144.1; --  
 KW Receptor.  
 SQ SEQUENCE 1294 AA; 141318 MW; 88658384859822A1 CRC64;

Local Match  
 Local Similarity 25.8%; Score 1593.5; DB 10; Length 1294;  
 Conservative 175; Mismatches 434; Indels 299; Gaps 42;

QY 5 SSEFFLSVTTLFFSFFSLSPFOSPSQSLREIHLQISFDVLPD-KNLLPDW-SNNKPC 62  
 DB 4 NSFWLPT---LVSFPIISAMAES---RDISTLPTLADSIIEGKGFRLNMFDSPTPC 55  
 QY 63 TFGVTCRDKKTSIDLSKPLAVGSAVSSLSITGLESLEFNSHINGSVGFKSA 122  
 DB 56 SWGICIGIHNVVAIDLSVPLAAPPICIGAFQSLV-----RNFSGCGP----- 101  
 QY 123 SLTSLDLSRNSLSGPVTTLSLSCGSLKFLNVSNTLDFPGVSGGLKLANLEVLDSA 182  
 DB 102 -----SGELP--EALGNLQNLQYLDLSNNELGPTPIIS-LVNLKMLKENVLDY 146  
 QY 183 NSISGANVGVWVLSDCGELKRI--LAISGNKISGV--DVSRCVNLEFLDVSSNNFSTG 237  
 DB 147 NSLSGQ-----LSPAIAGLOHLTKLISWNSSISGSLPDLGSKLLELDIKMTFNDS 200  
 QY 238 TP-FLDGCALQHLIDISGNTLSDPSRAISTCTELKLNINISNQFVGPPI----- 286  
 DB 201 IPATFNLSCGLHFDASQNNLTGSIPTGITSITNLLTLDSSNSFGTIPREIGOLENIE 260  
 QY 287 -----PLP-----LKSLOYLSLAENKFTGEIPDFLSGACDTLTGLDLSGNHFFYG 330

DB 261 LILGKNDLTGRIPQELISLQKLLHLEECQGTGKIPIWSISG--LSSITLIDLSNNPDA 319  
 QY 331 AVPPFGCSLLLESLALSNNPFGELPMDTLTKMRGLKVLVDLSFNSBSGELPESUTNLSA 390  
 DB 320 ELPSSMGEKGNLTOLIAKNAGLSGMPKE--LGNCCKLVINLSFNAIIGPIPEEPADLEA 378  
 QY 391 -----SLLTLDLSNNPFG--PILP----- 408  
 DB 379 IYSFVEKNGKLSGRVPDWIQKKAKARSIRLGQNNFSGGLPVLPIQHLISFAESNLSGS 438  
 QY 409 ---NLQNPKNLTLOELVYQNNGFTGKIPTLSNCELVSLHSFNYLSGTPSSLSISL 465  
 DB 439 IPSHICQ--ANSLHSLILHNNLTGTIDEAFKGTNLINLMDHNIHGEVPGVLAEL-P 495  
 QY 466 LDDKMLNMLEGHIPOELMYKTELETLIDPNDLTGEIPGSLNCTNLNMTISNNELT 525  
 DB 496 LVTELSQKFPAGMLPAELWESKTLLEISLSNNEITGPIPSIGLSLVQRLHDNNILE 555  
 QY 526 GEIPKWIQRLNMLALIKLNNNSFGNIPDELGDCRSLLIWLDTNLPNGTIPAMF--- 581  
 DB 556 GPVPSGVDRLNLTNLSLRGNLSGIIPLALFNCKRLATLDSYNNLTGNIPSAISHLTL 615  
 QY 582 -----KQSGKIAPAFIAGKRYVYIKNDGKKECHGAGNLEFQIGRSEQLRLSTR 632  
 DB 616 LDELSSNQLSGSIPIAICV-----GFENEAPDSEHFLQHGILDLSTYQLTGQ 665  
 QY 633 NPENITSRVYGGHTSPTPNNGSMFLDMSYNNLSGYIPEKISGMPVYFLINTGND--- 689  
 DB 666 IPTSI-----KNCAMVWNLNQGMLNGTIPVEIGELTNLTSLNSNEVVG 712  
 QY 690 -----ISGISPEVGD-LRGLNLTLDLSNNKLDGRIPQANSALTW 727  
 DB 713 PMLPWSGPIVQQLGILLNNHLDGSPAKIGQILPKIATLDSNALGTLPOSLLCNVY 772  
 QY 728 LTEIDLSNNLSGPI-----PEMQFETPPAKFLN 758  
 DB 773 LNHLDVSNHLSGHIQFSCPDKEYSSTLLFPNSSNHFSGSLDSEISNFTQLSTLIDHN 832  
 QY 759 NP-----GLCG-----YPLPR-- 769  
 DB 833 NSLTGRLPSALDLSLNLVLDLSNNLYAIPCGICNIFGLSPANFSGYIMYSLADA 892  
 QY 770 ---CDPSNADGYAHHQRSHGRPPASLAGSVAMGILFSFVCIIGLIVGREMKRR-RKK 824  
 DB 893 AGGICSTNGTDKHALPHYRVRRAITICA-----FTVILIVLAVLYRRKLVSRR 945  
 QY 825 EALEMYAEGHNSGDRYANNITW-KLTGVK--EALSTINLAPEKPKLKTFRADLIQATN 881  
 DB 946 PLAFE-----SASKAKATVEPTSTDELGKRSREPLISINLAFTEHALLVTDILKATE 1000  
 QY 882 GFINDLSIGSGGPDYKAILKDGSAVAIAIKLHVSG--QGDREFFAMMETIGIKERNL 939  
 DB 1001 NFSKWHIIGGGGTYKALPEGRVAILKRL-HGHQHQGDREFFAMMETIGIKERNL 1059  
 QY 940 VPLIGYGVGDERFLIYEWENGSLEMWLNRNADALEALGMPDLKICISARGAFLAHL 1119  
 DB 1060 VPLIGYGVGDERFLIYEWENGSLEMWLNRNADALEALGMPDLKICISARGAFLAHL 1119  
 QY 1000 NCSPIIHRDKSSNVLLENLEARYSDGMARLMSAMTHLSVSTLACTPGVPEPYQ 1059  
 DB 1120 GFVPHIHRDKSSNVLLENLEARYSDGMARLMSAMTHLSVSTLACTPGVPEPYQ 1178  
 QY 1060 SFRSGTGDDVYSYGVVLELLTGKPTDSDP--GDNNLVGVV--QAKKRIISVPEPELM 1117  
 DB 1179 TMSKSTGDDVYSYGVVLELLTGKPTDSDP--GDNNLVGVV--QAKKRIISVPEPELM 1117  
 QY 1118 KEDPALEI---ELLQHLKVAVACLDDRAMRPTMVOMAMFEK 1156  
 DB 1238 ---PVSSVWREQMARVLAIRDCIADBPFRKPTMLEVYKSLK 1276

RESULT 11  
 Q9XE13



ID 09XEL3 PRELIMINARY; PRT; 1420 AA.

AC 09XEL3:

DT 01-NOV-1999 (Tremblrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DE EST C96716(C10608) corresponds to a region of the predicted gene.

OC Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Euphorbiaceae; Oryzaeae; Oryza.

CC NCBI\_TaxID=4530;

CC [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, NIPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa Nipponbare (GAS) genomic DNA, chromosome 2, PAC

RT clone: P0437H03 (contig b).";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

CC -1-SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AP000367; BAA82393.1; -

DR InterPro; IPR000719; Euk\_pkinase.

DR InterPro; IPR001611; LRR.

DR InterPro; IPR003592; LRR\_out.

DR InterPro; IPR002290; Ser\_thr\_pkinase.

DR InterPro; IPR004040; STY\_pkinase.

DR InterPro; IPR001245; Tyr\_pkinase.

DR Pfam; PF00560; LRR; 30.

DR Pfam; PF00069; pkinase; 1.

DR PRINTS; PRO0019; LEURICHRPT.

DR PRINTS; PRO0109; TYRKINASE.

DR ProDom; PD000001; Euk\_pkinase; 1.

DR SMART; SM00370; LRR; 24.

DR SMART; SM00221; STYK; 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.

DR ATP-binding; Serine/threonine-protein kinase; Transferrase.

SQ SEQUENCE 1420 AA; 153401 MW; B938CE5B9AE92F5 CRC64;

Query Match 24.3%; Score 1503; DB 10; Length 1420;

Best Local Similarity 32.0%; Pred. No. 2,5e-89;

Matches 439; Conservative 186; Mismatches 462; Indels 284; Gaps 40;

QY 6 SFPLSVT-ILFPPSPFSLFOAPSQSLYR-RIHQILSP-KQVLDKDKLIDPMSNKK-P 61

DB 99 SFIVPVSKPAHYHCLPTLLCFIPITALAESITKULFAIKAIYAGKPELHNWELTEPP 158

QY 62 CTPEGVTCDDDKTYSIDLSKPLANGFSVSSLSLTGLSESLFLSNHINGSVGFCS 121

DB 159 CMMSSGISCGLTVVVAIDLSSTPLVDF--PQIITAFOSLVRLINVSQCGFSGELPEAMVN 215

QY 122 AS-LTSLDLSRNSLSGPVTVTLTSLSCSGKIKLVNSNTLDPPKVGSGL-KLNSLEYLD 179

DB 216 LQHQLHLDLSNDQGGFLP--ASLPDLKMLKVMVLDNM--FSQQLSPALNHLQQLTVLS 271

QY 180 LANSISGANVGVWVLSQCGELKHLAISGNKISGDVVSRCVNLFEFVDSNNFSGIP 229

DB 272 ISTNFFSGG-----LPELISGLK-----NLEIYDITNAFSSGIP 306

QY 240 -FLGPCSLQHLDISGNLSDGFSALSTCTELKILNINSQVGPFP--LPLKSLQYL 236

DB 307 ASFSVLRLILYDANNNNLTSIPGIRALVNLVLDLSSNGVIGALPKELCQKLNQSL 366

QY 297 SLAENKFTGEIPDFLS-----GACDVLTLGLDSGNH----- 327

DB 367 ILSDNELTGSIPBEIIGNKOLEVNLKCNLMQTVPISTIGLETLBELGLYSFNSFSGELP 426

QY 338 -----FYAVPPFFGSCSLLESLSLSSNNFSGELPMDTLKMGELKY 369

DB 427 ASVGEILNLRLQMLMAKSAGFTSIPKEIKNCCKLTTVLISGNFTGTRP-BELADIVAVVL 485

QY 370 LLSFENFSGELPE----- 383

DB 486 PDVSGNRLSGHIIPDWIOWNSVSSISLAQNMFDGPPGLPLHLVFSFSAESNRLSGIPAK 545

QY 384 -----SLTNIS-----ASLTLLDS 398

DB 546 ICGFTLOILRLANDNNLTGSIDETFKCKKRLTSLDNLHNGEIPRYALLPLVSLDS 605

QY 399 SNNFSGPLPLCONPKNTQELYLQNGGTGKIPPTLSNCSSELYSLHSFNYLSTGPS 458

DB 606 HNNFTGIPDLMS--STLLDLSNDQLTGWTESIKGLSLQSLSIDRYVLQGPLR 663

QY 459 SLGSLSKLRDLKMLNMLEGEPQELMYVYKTELTLLFDENDLTGELPSGLSNCTUNMIS 518

DB 664 SIGALRLVTLTSLSGNNLSIEDIPQLFNCRNLVTLLSCNNLTGHIIPALISHTLNTLV 723

QY 519 LSNRLTGEIP-----KIGLENTALIKLSNNSFGNIPDELGDCRSLLMWD 566

DB 724 LSRRLSGALPSBELCVAFSRESHSELEYQHIGLIDLSRNLRTGHIIPRAINNSLIVELH 783

QY 567 LNTVLFNGTIPAMFKSG--KIANFLAKRYVYIKNDGKKEGAGNLLEPGIR 622

DB 784 LQNLISGTTIPVELAEIRNITITLSSNALVGPVLP-----PVPLASLQGLL 831

QY 623 SEQNLRLSTNPNITSRVYGGHTSPEDNNGMMFLMSYNNLSGYIPKEI----- 674

DB 832 LSN-NRLSGIPSGI-----GNILP-----QIMLDSGNALTGTLPLDLLCKESLNH 878

QY 675 -----GSMFY-----LFLNLGNDISGIPDEVGDLRLNLIDLSNKL 714

DB 879 LDVSDNNISGQIPFSCHEDKESPIPLIFPNASNNFSGSLDESISNFKLYLIDLHNLSL 938

QY 715 DGRIPQMSALTMLTEIDLSNNLSGPFP--EMGOFE--TEPPAFINNPGLCGPLPRCD 771

DB 939 TGRPLPALARVTSIYLDLSSNDPFGTIPCGICMGFLTP--ANFSGRDGRTPLDCA 996

QY 772 PSNADYVHHQ--RSHGRPPSLAGSVMGLLFSFVCFGLILVREMKRRKKEALE 829

DB 997 AEEGVCANRRVDRKPPHFEVLEATICIAITAVILVVLVYLRKRRKMLRRQFV 1056

QY 830 MYAEGHNGSGDFTANNMKLTG-----VKALSLINLAEPKPLKLTPLADLQATNGFH 884

DB 1057 LVPAGDNMADEHETTLN-NLGRMRKRRPPSINLATEFHAVRVTVBEIMRATNGFD 1115

QY 885 NDSLISGGFGDVYKAIKDSAAVAILKLVHSGO-----GDREPMAMETIGIKKHNV 940

DB 1116 GNAVVGDFGFGTVYRAELPGRRAVAVKRL-HGVRRKFGGGERFRAMEVGVKVRHNV 1174

QY 941 PLLGVCXGDERLLVNEVWKGLSEDLVLDPKKGVKULKSTRKIALISARGIAFLHN 1000

DB 1175 PLLGVCAGADRFLVYETMEHSLIEDRLRG--GGGALGWPERLTICGGAARGLAFLHNG 1232

QY 1001 CSPHILHRDMKSNVLIDENLEAVSPFGMARLMSANDTLSTAGTGYPPRYGOS 1060

DB 1233 FVPHVHHDVSSNVLLEGGLQPRVSPGLARITISACEHVS--TVLAGTGLGYIPREYALA 1291

QY 1061 FRCTGKGVSYGVVLLLELTGKRP-----DSPDFGNNLVGWK--QHALLRIS 1109

DB 1232 MKCTAKGVSYFVGVMLELLTGRPTWSABVYTABGDDEGGGSLGVKRMMAARGGG 1351

QY 1110 DVPEBELMKEDPALEIELLOHAKYAVACLDPRAMRPRTYQVAMPEKIQ 1160

DB 1352 EYFDACLPVSGARE--OMARVLDVADCTADEPWRPRTYAVARVGAITEA 1401

RESULT 12

Q9SN91

ID 09SN91 PRELIMINARY; PRT; 1232 AA.

AC 09SN91:

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE Leucine rich repeat-like protein.

GN FIC12.60 OR ATG20140.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Terry N., Ardiles W., Buyschaert C., Dasseville R.,  
 RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,  
 RA Villarejo R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,  
 RA Mayer K.F.X., Lemcke K., Schueller C.,  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-305 FROM N.A.  
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Terry N., Ardiles W., Buyschaert C., Dasseville R., De Clerck R.,  
 RA De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarejo R.,  
 RA Gielens J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AL022224; CAA18239.1; -  
 DR EMBL: AL161552; CAB79014.1; -  
 DR HSSP: P08631; 1AD5.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR\_out.  
 DR InterPro: IPR003592; LRR\_out.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00560; LRR; 30.  
 DR Pfam: PF00069; Kinase; 1.  
 DR PRINTS: PRO0019; LEURICHRPT.  
 DR PRINTS: PRO0109; TYRKINASE.  
 DR PRODOM: PD000001; Euk\_pkinase; 1.  
 DR SMART: SMO0370; LRR; 30.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; UNKNOWN\_1.  
 DR PROSITE: PS0011; PROTEIN KINASE DOM; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR ATP-binding, Serine/threonine-protein kinase; Transferase.  
 KX [1]  
 KX SEQUENCE 1232 AA; 135534 MW; EAD6322C8B967726 CRC64;  
 Query Match 23.0%; Score 1419; DB 10; Length 1232;  
 Best Local Similarity 31.2%; Pred. No. 6.2e-84;  
 Matches 402; Conservative 221; Mismatches 441; Indels 224; Gaps 41;  
 QY 9 LSVTLFFSFSFSLGSPQSPQSLVREIHQLSFKDVL-----PDKNLLPDMS-NKPNCT 63  
 DB 4 LVLLLFILCFSGSL--GPG-IINNDLOTLEVKKSLVTNPQEDPLRQNSNDINVC 59  
 QY 64 PDGVTC-----RDKVTSIDLSKSPINVGSAVSSLL 96  
 DB 60 WIGVTCNDNGLRVYALNLTGLTGISIPWFGFNDLTHLDLSNNL-VG--PIPTALS 116  
 QY 97 SLTGESLFLSNHINGSVG-----FKCS 121  
 DB 117 NLTSLSLFLFSNQLTGEIPSQGLSVNRSIRIGDNELVGIPTTLGNLVNLQMLALS 176  
 QY 122 ASLT-----SLDLSRNLGSPVTLTSLGSCGKELINVSNTLDPGKYS 167  
 DB 177 CRTLGPSPQLGLVAVQSLILQDNYLGGPIP--ALGNCSDLTVTAAEMNLN--GTIP 232  
 QY 168 GGL-KINSLSEVLDSLANSISGANVGVSLDGGC--LKHAIISNKTISGDV--DVSTC 221  
 DB 233 AELGRLENLEILNANSLTGE-----IPSQGEMSOLOLYSLMANOGLIPKSLADL 286

QY 222 VNIIEFDVSSNPNFSTGIP-FLGDCGALQHLDISGNKLSGDSFRAI-STCTEIKLINISSN 279  
 DB 287 GNLQTLDSLANNLTGRIPEEFNMNSQLDLVIANHNLGSLPKSICSNNTNLEQLVSGT 346  
 QY 280 QFVGPIIP--LPLKSLQVSLAENKFTGEIPDPLSGACDTLTGLDSGNHFGYAVPPFG 337  
 DB 347 QLSGEIPVLSKQSLKQDLDSNLSAGSIPEALFEIYE-LTDLVHNNTLTGTLSPSIS 405  
 QY 338 SCGLSLSLSSNNFSGEIPMDTLKRGKLVLDLSNFSGEPESLNLASLITL 397  
 DB 406 NLNLMQVLYVHNHLEKLPKE-ISAIRLEVLFLYENRFSGEIPQEIENCT-SLKMDIM 463  
 QY 398 SNNFSGPILNLCQPKWTLOELVYNNNGFTGKIPPTLSNGSEVSLHSPNYLSGTP 457  
 DB 464 FGNHFESEIPPS--GRKELNHLRQNEIVGCLPASGCHQNIILADNQLSGSIP 521  
 QY 458 SSLGSLSKLDLKLWNLMEGEIPQELMYVKTLETILDPNDLTG----- 502  
 DB 522 SSFGLGLGLQMLVNNSLQNLPSLISIRNLTIRLNSHNLNGTILHPLCGSSSYLSD 581  
 QY 503 -----EIPSGLSNCTNLNWLISLNNRLTGEIPKWLGRLENLAIKLSSNFSGNIPD 554  
 DB 582 VTNNGFDEEIPLELGNQNLDRILKKNQUTGKIPWTIKIRLSLIDMSNALTGTP 641  
 QY 555 ELGDCRSILWDLNLTNLPNGTIPAMFK--QSGKIANTFAGKRYIYIKDGKKEGCHA 612  
 DB 642 QVLCCKLTHLDLNNNLSGIPSPWLGKLSQGLBK--LSNQFV--ESLPTLEFNC 694  
 QY 613 GNLL-----EFGQIRSEQLNRLSTRPCNITSRYVGHTSPFDNNGSMFLDMSYN 665  
 DB 695 TKLVLLSDGNSLNGSIPOEIGNGALVNLIDNQSGLPQAMGKSLKYLRLSRNS 754  
 QY 666 LSGVLPKEIGSMPL-PLINGNHNDISGIPDEVGDLRGNIILDSNKL DGRIPQMSA 724  
 DB 755 LTGEIPVEIGQLDQSLADSYNNFTGDIPTGTLSKETLTDLSHNQLTGEVYSGVD 814  
 QY 725 LTMLEIDLSSNNLSGPIPEMGQFETPPPAFLNPPGLCYPLPRCPSPNADGYAHHQR 784  
 DB 815 MSLQYLVNVSFNNLGKLLK--QFSRWPADSFLLNTGTCSPSLRCN----- 859  
 QY 785 HGRPPASLAGVAMGLFSPFCIFGLIVGEMRRRR--KKEALEMYAEGHNSGDR 841  
 DB 860 ---RVRTISALPALEMLVIALP-----FKORHDFKK-----VGHGSTAYT 899  
 QY 842 TANNNNWKLTVKEALINLAFAEKPLK-----LTFADLLQATNGFHNDSLISGSGF 894  
 DB 900 SSSSSS-----QATHKPLFRNGASKSDIRMEDIMEATHNLSSEFMISGSGS 945  
 QY 895 GDVYKAILKQSAVAIKLIVSG--QGDREPAEMETGCKIKNRYLPLGYCKVGD-- 951  
 DB 946 GKYYKALENGEIVAVKTLIMDDLSNKSFSREYKVTGRIHRHVLVKLMGVCSSKSG 1005  
 QY 952 RLLVNEVMKYGSLIEDLQDPK---KGVYKLSLTPRKAIQISARGLAFLHNHNSPHI 1007  
 DB 1006 NLLIYEMKNGSITWDLHEDKYLEKKKLLDMEARLRAVAGLAQVETLHDDCVPIVH 1065  
 QY 1008 RDMKSSNVLLDENLLEARVSDFGMARLMSA-MDTHLSVT-LAGTGYVPEPYGFRSCT 1065  
 DB 1066 RDKSSNVLLDSMEALHDFGLAKVLTENCNTDNTDSNWFCSYGYLAPEVAYSLKATE 1125  
 QY 1066 KGDVYSYGVVLLLELTKGRPTSPDGDNNLVGMVYQKHAELIS---DVFPDELKKEP 1122  
 DB 1126 KSDVYSWGIIVLMEIVTKRPTDSVFGAEDMVRMVAETHLEVAGSARDKLDIKLAPL 1185  
 QY 1123 LATELLOHLKVAACLDDBRAMRRTVWQ 1150  
 DB 1186 EEDACQVLEIALQCTKTSQGERPSSRO 1213  
 RESULT 13  
 Q9FN37 PRELIMINARY; PRT; 1036 AA.  
 ID Q9FN37



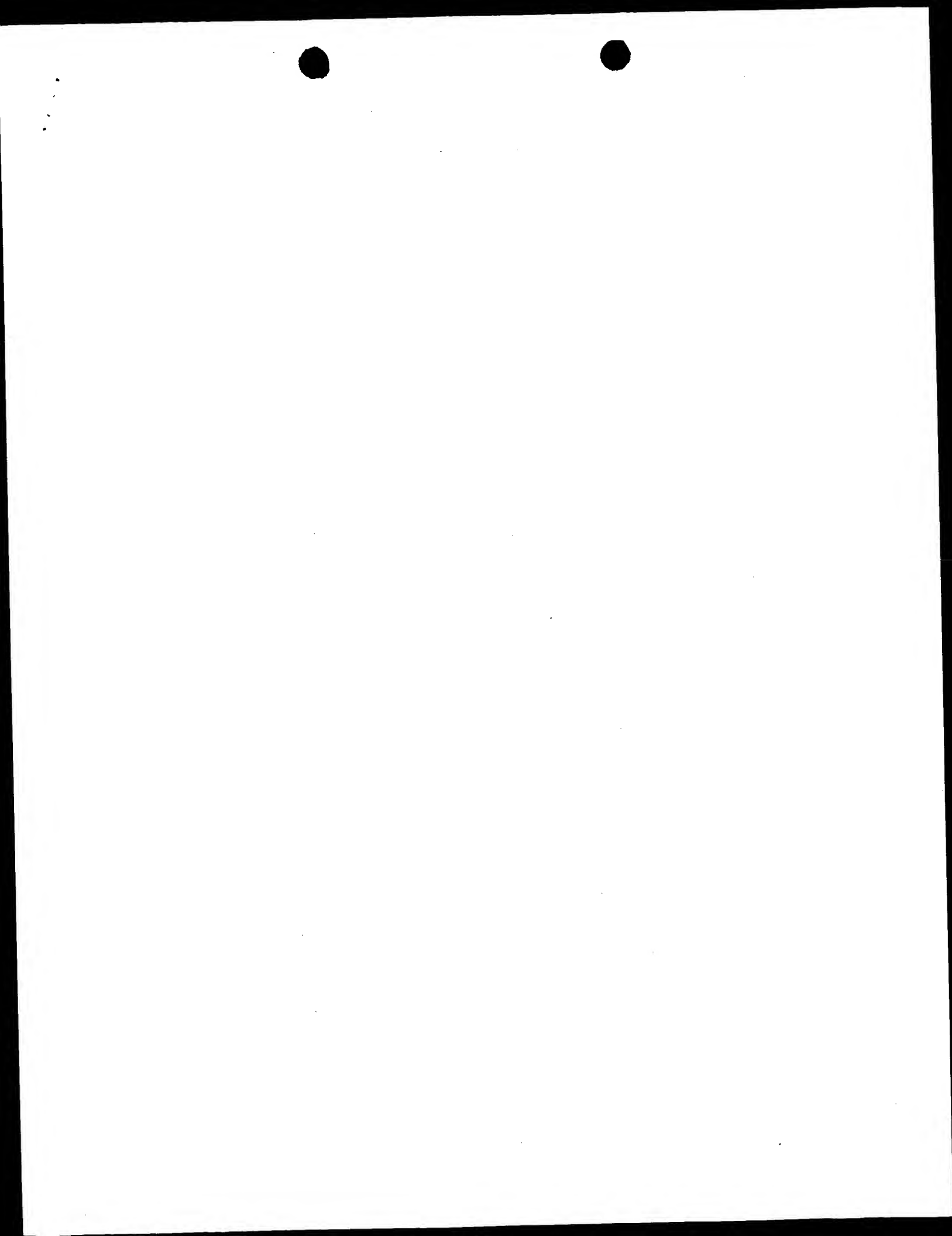
QY 1060 SFRCTKGDVSYGVVLELTGKRPD-SPDFGDNVGVWQ-HAKLRISVDPELM 1117  
 DB 930 SIAACRGVVSFGVLELVGRFVPCVKCKSRDLSRVFQWAKREABLDITTI- 988  
 QY 1118 KEDPALEIELQHLKVAACDDRAMRPTVQVAMKEI 1158  
 DB 989 -RENNERTVLEMLIACKCIDHEPRRRLIEVVTLWEDL 1028  
 RESULT 14  
 ID 09LVPO PRELIMINARY; PRT; 1102 AA.  
 AC 09LVPO;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Receptor-like protein kinase.  
 DE Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OA NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=20181125; PubMed=10718197;  
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
 RA Tabata S.,  
 RA Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
 RT clones.";  
 RL DNA Res. 7:31-63(2000).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AB019227; BAA96896.1; -;  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR01245; Tyr\_kinase.  
 DR Pfam: PF00560; LRR; 23.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR PRINTS: PRO0019; LEURICRPT.  
 DR PRINTS: PRO0109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR SMART: SM00370; LRR; 22.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 YR [REDACTED]  
 SEQUENCE 1102 AA, 120476 MW, 6CDP852ADD3D54D CRC64;  
 Query Match 22.3%; Score 1380.5; DB 10; Length 1102;  
 Best Local Similarity 32.0%; Pred. No. 1.7e-81;  
 Matches 387; Conservative 179; Mismatches 470; Indels 173; Gaps 31;

QY 271 LKLNISNOFVPIPLP--PKSLQVLSLENKFTGEIPFLSGACDTLTGLDSGNH 328  
 DB 267 LFTLALYKNQVLPVPELGDLOSLEFLYLVNGANGNTPRI -GNLSVALEIDFSENAL 325  
 QY 329 YGAVPPFPGSCSLLESIALSSNNFSGELPMDTLTKMRGLKVLDSFNEFSGELPSLNTL 388  
 DB 326 TGEIPELENGIELELVLFENQLTGTIPEV -LSTLKNLSKLDSINMALTGPIPLGFOYL 384  
 QY 389 SASLLTLDSSNNFSGPIPLNLCQNPKNLQELVYLNNGFTGKIPPTLSNCELSVLSLS 448  
 DB 385 -RGLFMLQLEFQNSLGTIPPKL--GWYSDLVWDSMDNLSGRIPSYLCIHNMTIINLG 441  
 QY 449 FNYLSGTPESLISKLKDLKMLNMEGEIPEELMYVKTLETLILDNDITGEIPSEL 508  
 DB 442 TNNLSGTPIGITCTKTVGLRLARRNNLVGRFNSNLCQVNVATLELGNRRRGSIPREV 501  
 QY 509 SNETNLWISLNNRLTGEIPKMTIGLENLALIKLSNNSFSGNIPPELDGCSLWLDN 568  
 DB 502 GNCBALQRLQADNGFGEIPREIGMLSQGLTINISSNKLITGEVPEIRNCKLQRLDWC 561  
 QY 569 TNLFNGTIPAMFQSGKIAMNFIAGKRYVYIKNDGKKKECHGAGNLEFQIRSQLR 628  
 DB 562 CNNSGTLPSSEV-----GSLYQL-----ELLK 583  
 QY 629 LSTRNPNITSRVYGGHTSPFDNNGSMFLDSYNMLSGYIPEKLSMPLFI -LNLGH 687  
 DB 584 LSNNN-----LSGTIPVALGNLSRLTELQWGNLFNSGIPRELSLGLQIALNLST 635  
 QY 668 NDIGSIPDEVGDLRGANILDSNKLKGRIPOMASALTM/TELDSSNNLSGPIPEMG 747  
 DB 636 NKLITGEIPPELSNLMVLEFLLNNNNLSGIPSSFAULSLGTFNPSNLTGPIPL-- 693  
 QY 748 PETPPAKFLNPNPGICGYPLPRC-----DESNADGYAHQSHGRPPASLAVAMGL 801  
 DB 694 -RNLSMSFIGNBLCGPLNQCIQTQPPASQSTG-----XPGGRSKIIAIT 742  
 QY 802 PSFVCIPLGLIVGRMRKRKKAELEMYAEGHNGSGDRANTNKKLGVKALSNL 861  
 DB 743 AAVIGVSLMLALIVLMRRPVRVTVASSAODG-----QPSMSL 782  
 QY 862 AAFEPKRLKTFPADLQATNGFHNDSLIGSGFGDVYKALIKDSAVAIKLI--HVSQ 919  
 DB 783 DIFPPKSGFTQDVAATNDPDESFFVGRGACGVYAVLPAGYTLAVKKLASNHEGN 842  
 QY 920 G---DREFMAEETICKIKNRLVPLCYGVYDERLLVNEVMYKGSLEDTLQPKKGV 976  
 DB 843 NNNVDSFPALITLIGNIRHRNIVKLHGFCNHQSGNLLLEYMPKSGSLGILHDP--SC 899  
 QY 977 KKLSTRKIIAGSAGLAFLHNCSPHIIRDMKSNVLDENLEARVSPFGMARLMSA 1036  
 DB 900 NLDWSKRFKIALGAQGLAVLHDDCKPRLFRHDIKSNILLDDKREAVVGGFLAKVID- 958  
 QY 1037 MDTHLSVSTLAPGVPVPEYQSPFCSCTKGDVSYGVVLELTGKRPDSDPDGNNL 1096  
 DB 959 MPRSKMSALASGYIAEYAVYMKVTEKSDIYSGVVLLELTGKRPVQPIDG-GDV 1017  
 QY 1097 VGVYKQHAQ--LRISVPPELMKEDPALEIELQHLKVAACDDRAMRPTVQVAM 1154  
 DB 1018 VNNVRSYTRDALSGVLDARLTLEDERIVSHMLTVLKTALCTSVSPARESMQVVM 1077  
 QY 1155 FKEIQAGSG 1163  
 DB 1078 LIESERSEG 1086  
 RESULT 15  
 ID 09CA77 PRELIMINARY; PRT; 1079 AA.  
 AC 09CA77;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)

DE Putative receptor protein kinase.  
 GN FLM20.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eucaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CY. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federpiet N.A., Kaul S.,  
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.V.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Eggen P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,  
 RA Miltseher J., Miranda M., Nguyen M., Nielsen W.C., Osborne B.I.,  
 RA Pail G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shum P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Taitum M.J., Town C.D., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter U.C., Davis R.W.,  
 RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana";  
 RL Nature 408:816-820(2000).  
 DR EMBL, AC011865; AAC52362.1;  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003591; LRR\_Tyr.  
 DR InterPro: IPR002290; Ser\_Thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00560; LRR; 17.  
 DR Pfam: PF00069; Kinase; 1.  
 DR PRINTS, PR00019; LEURICHRPT.  
 DR PRINTS, PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART, SM00369; LRR\_TYR; 7.  
 DR SMART, SM00220; S\_TKc; 1.  
 DR SMART, SM00219; TYKc; 1.  
 DR PROSITE, PS00107; PROTEIN KINASE ATP; UNKNOWN\_1.  
 DR PROSITE, PS00101; PROTEIN KINASE DOM; 1.  
 DR PROSITE, PS00108; PROTEIN KINASE ST; 1.  
 DR KIP, ATP-binding; Kinase; Receptor; Transferase.  
 S SEQUENCE 1079 AA; 118675 MW; A19918DD9B8D44 CRC64;  
 Query Match 22.3%; Score 1380; DB 10; Length 1079;  
 Best Local Similarity 32.2%; Pred. No. 1,8e-81;  
 Matches 392; Conservative 204; Mismatches 406; Indels 214; Gaps 44;  
 QY 26 ASFSGLYEIRHQLISFKVLPDKN-----LLPDWSNKNP--CTPDGVC--RDDKVTYS 76  
 DB 5 AVAGDLSDREVLTLKSLKSLRNPNQGLYTEWKNQDVQWCPGIICTPGQSRVYG 64  
 QY 77 IDS-----SKPLNVGSASVSSLSLTGLSLFLSNHINSVSGFKCSASLTLDLSRN 132  
 DB 65 INLTSTTIGPFLKVPKFSALT-----ELTLDLSRN 94  
 QY 133 SLSGPVTLTSLGSCSGKLTFLVSSNTLD-----PPKVSGLKLSLEVLDSANSISGA 188  
 DB 95 TIEGRIP--DLISRCNHLNHLNLSNLTLEGLSLPG-----LSNLEVLDSLNRITGD 145  
 QY 189 NVGVAVLSDGCGELKHLAISGNKISGDV--VSRGVNLEFLDVSSNFS-----TGI----- 238  
 DB 146 IOSGFPLF--CNLSIVANLSTNNFTGRIDTLFNGCRNLKLVYDSSNPSFSGWVTFGRGLV 203  
 QY 239 -----PLGDCSALQHLDISGNKLSGDFSRALSTCTELKILNINSNOFY 282  
 DB 204 ERSVADNHLGNISASWFRNGCT--LQMLDISGNAFGGEFPGQVSNCONLNVLMKNKFT 262

QY 283 GPIPL--PLKSLQVLSLAENKPTGEIPDLFSGADTITLGDLSGNHFYGVPPFGSCS 340  
 DB 263 GNIPARIGISSLSKGLYLGNNTFSRDIPELTMLTN--LVFIDLSENKGGIOIEIFGRFT 321  
 QY 341 LLESLALSNNSGSELPMDTLLKRGKLVLDLSENFSEGESPELNLNLSASLTLDLSN 400  
 DB 322 QVYVLTANSVYGGINSNLIKPNLSRDLGYNNFSGQLPTEISQI----- 369  
 QY 401 NFGSILPMLCQNPNTLOELYLQNNGTGKIPTPLNSCSLSVLSHLSFNYSGLTIPSL 460  
 DB 370 -----QSLKFLTLAVNNSGDIPOEYGMPELQALDLSFNNLTGSIASF 414  
 QY 461 GSLSKRLDKLMLNLEGEIPQELMYVTLLETLILDFNDLGEIPSGLSNCTNLNLSLS 520  
 DB 415 GYLTSI-----LW-----LMLNNSLSGELPREIGCTSLMFWVA 450  
 QY 521 NNRLGEIPKMTIGRL--ENLALILKSNNSPSGNIPDELGDGRSL--IWLDTNLTNFGTIPA 578  
 DB 451 NNQLSGRHPPELTHRGSNPSPTFEVNRQNKILAGSGEGLAMKRWIPARFPFN--FYVA 509  
 QY 579 AMFKSGKIAANFLAGKRYVYIKDGAKKKECHGAGNLLEFGIRSEQLRLSTRPCNT 638  
 DB 510 ILTKKSGSLMDHY-----LKGYGLFVVC--SAGST-----VRLKISAY-----LOLS 551  
 QY 639 SRVYGHSPFTFDDNNGSMPLDMSYNNLSGYIPKEIGSMPEYFLIMLGHNDISGSIPEDE 698  
 DB 552 GNKFSGEIPASISQMDRLSTLHLGFNEFEGLPPEIGQLPLAF--LNLTNNSGELIPOEI 610  
 QY 699 GDLRLNLTLDLSNKLDRILPOAMSLTMTLELDLSNN--LSGPIPEWQFETPPAKEL 757  
 DB 611 GNLKCLQNLDSLTFNNSGNFTSLINDLSELKSNISYNPISGALPTTQVATFPRKDFL 670  
 QY 758 NNPLCGEPLPRCDPSNADGYAHRSH--GRRPASLA--GSVAMGLLPSFVCIPLDI 811  
 DB 671 GNP-LRFP-----SFFQSGNNTKISNOVLGNRPRTLLIISLALAF-----LACLV 721  
 QY 812 LVGRMKRRKRKEAELEMYABEGHNSGRTAN--NTNWKLTGYKALSLNLAPEKPL 868  
 DB 722 VSGIVLVNVAASREABIDL--LQSKTRHDMTSSGSGSSFW--LSGKIKYIRLDKSTF----- 775  
 QY 869 RKLTPADLQATNGFNHDSLISGGGFDVYKAILKDGSAVAILKLIHVSQGDDEFAEM 928  
 DB 776 ---TVADILKATSNFSEKRVVGRGVYIRGVLPDGEKVAKQLQREGTAERKEFRAEM 832  
 QY 929 ETI-----GKIKRNLVPLIGCTKYGDERLLVNEVMKXGSLDELVDPKKGVKIKLSTR 983  
 DB 833 EUSANAFEDMAHPMLVRLYGMCLDSEKILVHEYMGGSLLEELITD-----KITLQWK 886  
 QY 984 RKIATGS--ARGLAFLEHNCSPHIIHRMKSSNLTLDENLEARYSDGMARLNSAMDTHL 1041  
 DB 887 KRDIATDVARGLVFLHIECPYSIYHRVYKASNVLLDGHGAARVATDGLARLLNVGDSHV 946  
 QY 1042 SVSTLAGTPGYTPPYVYVFCSTKGDVYSGVVLLELTGKRPDPSDPGDNNLVGMWK 1101  
 DB 947 S-TVLAGTGYVAPYEGTQWATTRGDVYSGVLTMLTAGRAVADG--GEGLVEMAR 1002  
 QY 1102 QHAKRLISDVDPPELM--KEDPALTEILQHLKVAVACLDPAAMRRPTVQVAMFKEI 1158  
 DB 1003 RYMTGNMTRAKSPITLSGTQKGNAG--QMTLELKIGVCTADHPARPNMKEVATLAVIKI 1061  
 QY 1159 ----OAGSINDOSTI 1170  
 DB 1062 SGKLELFWLSSOGYI 1077

Search completed: March 10, 2003, 18:18:41  
 Job time : 63.1415 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:16:35 ; Search time 19.0143 Seconds

(without alignments)  
2608.861 Million cell updates/sec

Title: US-09-823-394-2

Sequence: 1 MTFSSFFLSVTLTFFSF.....GFSTLMDMSIKVEPEGL 1196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1311	21.2	999	RLKS_ARATH	P47735 arabidopsis
2	1225.5	19.8	980	Q98YQ8 arabidopsis	Q98YQ8 arabidopsis
3	834	13.5	942	TMKL_ARATH	P43298 arabidopsis
4	639	10.3	674	TMKL_ARATH	P33543 arabidopsis
5	543.5	8.8	389	NAK_ARATH	P43293 arabidopsis
6	528	8.5	901	CR14_MAIZE	O24585 zea mays (m
7	508	8.2	410	APRA_ARATH	O06548 arabidopsis
8	507.5	8.2	412	APRA_ARATH	P46573 arabidopsis
9	507	8.2	817	KPRO_MAIZE	P17801 zea mays (m
10	498	8.1	849	SRK6_BRAOL	O09092 brassica ol
11	401.5	6.5	372	D100_ARATH	O00874 arabidopsis
12	393.5	6.4	582	SHO2_HUMAN	Q9uq13 homo sapien
13	392	6.3	710	IRAI_MOUSE	O06406 mus musculu
14	384.5	6.2	501	KEPL_DROME	O06552 drosophila
15	378.5	6.1	582	SHO2_MOUSE	O08520 mus musculu
16	378	6.1	496	TMW_ARATH	O098d1 arabidopsis
17	377.5	6.1	712	IRAI_HUMAN	P5617 homo sapien
18	341.5	5.5	821	CR1_ARATH	O05609 arabidopsis
19	338.5	5.4	1584	KTKI_DICDI	P24466 ascarisomye
20	331.5	5.3	1315	CHAO_DROME	P12024 drosophila
21	329.5	5.3	2300	CYAA_NEUCR	O01636 neurospora
22	328	5.3	342	PG11_PHAVU	P33334 phaseolus v
23	327	5.3	342	PG13_PHAVU	P58823 phaseolus v
24	324	5.2	342	PG12_PHAVU	O01513 podopora a
25	323	5.2	342	PG11_PHAVU	O05513 podopora a
26	320.5	5.2	2145	CYAA_PODAN	O05513 podopora a
27	303.5	4.9	330	PG11_ARATH	O05513 podopora a
28	300	4.9	330	PG12_ARATH	O05513 podopora a
29	297.5	4.8	330	PG12_ARATH	O05513 podopora a
30	296	4.8	2493	CYAA_USTMA	O05513 podopora a
31	292	4.7	1692	BRAT_CHICK	O05513 podopora a
32	290.5	4.7	1692	CYAA_CHICK	O05513 podopora a
33	290	4.7	825	TRKC_PIG	P24786 sus scrofa

34	289	4.7	1002	1	EPB5_CHICK	O07497 gallus gall
35	288	4.7	977	1	EPB2_MOUSE	O03145 mus musculu
36	287.5	4.6	984	1	EPB1_HUMAN	P54762 homo sapien
37	287.5	4.6	985	1	EPB4_XENLA	Q91571 xenopus lae
38	287	4.6	902	1	EPB3_XENLA	Q91736 xenopus lae
39	286.5	4.6	976	1	EPB1_HUMAN	P21709 homo sapien
40	286.5	4.6	984	1	EPB1_RAT	P09759 rattus norv
41	286	4.6	832	1	ANR3_HUMAN	P57078 homo sapien
42	285.5	4.6	821	1	TRKB_MOUSE	P15209 mus musculu
43	285.5	4.6	821	1	TRKB_RAT	O63604 rattus norv
44	284.5	4.6	794	1	TRK1_LYMS	O76997 lymnaea sta
45	284.5	4.6	1024	1	POPC_RALSO	Q9rbs2 ralsomonia s

## ALIGNMENTS

RESULT 1	ID	RLKS_ARATH	STANDARD;	PRT;	999 AA.
AC	P47735;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Receptor-like protein kinase 5 precursor (EC 2.7.1.-).				
GN	RLKS OR AT4G28490 OR F2109.180.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
CC	eurosid 11; Brassicales; Brassicaceae; Arabidopsis.				
OK	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=94035150; PubMed=8220453;				
RT	Walker J.C.;				
RL	"Receptor-like protein kinase genes of Arabidopsis thaliana.";				
RN	Plant J. 3:451-456(1993).				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=20083488; PubMed=10617198;				
RA	Mayer K.F.X., Schueler C., Wambert R., Murphy G., Voickaert G.,				
RA	Poll T., Duesterhoeft A., Stiekema W., Ehtian K.-D., Terry N.,				
RA	Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,				
RA	Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,				
RA	Kreis M., Delany M., Pugsdomech P., Watson M., Schmidheini T.,				
RA	Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,				
RA	Vos P., Hohnselt U., Zimmermann W., Wedler H., Ridley P.,				
RA	Langham S.-A., McCullagh B., Bilham L., Robben J.,				
RA	Van der Schuren J., Grymoprez B., Chuang Y.-J., Vandenbusche F.,				
RA	Braeken M., Welljens I., Voet M., Bastiaens I., Aert R., Defoor E.,				
RA	Witzenecker T., Botte G., Ramperger U., Hilbert H., Brian W.,				
RA	Holzer E., Brandt A., Peters S., van Staveren M., Dirxse W.,				
RA	Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,				
RA	Bernlaiser S., Hempel S., Feldpausch M., Lambrecht S., Van den Daele H.,				
RA	De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,				
RA	Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,				
RA	Clark L., Doggett J., Hall S., Kay M., Lennard N., Mcay K., Mayes R.,				
RA	Pettit A., Rajandream M.A., Lyne M., Benes V., Rehnman T.-H.,				
RA	Borkova D., Bloecher H., Scharfe M., Grimm M., Loehnert S.,				
RA	Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs M., Fartmann B., Granderath K., Danner D., Herzl A.,				
RA	Neumann S., Argitrou A., Vitale L., Liguori R., Piravandi E.,				
RA	Messner O., Ouglet F., Clabaud G., Mendenhall A., Falber R.,				
RA	Schnabl S., Hiller R., Schmidt W., Lechany A., Aubourg S.,				
RA	Chefor F., Cooke R., Berger C., Montfort A., Casacuberta E.,				
RA	Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,				
RA	Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,				
RA	Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,				
RA	Frismann P., Haase D., Lemcke K., Mewes H.-W., Stocker S.,				
RA	Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,				
RA	Parnell L., Dedha N., Gnoj L., Schutz K., Huang E., Spiegel L.,				

RA Sektion M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stronking T., Kallio J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan A., Jones K., Dione K., Cotton M., Joseph C.,  
 RA Du H., Ali J., Bergtholt A., Jones K., Sun H., Lamar B., Jordan C.,  
 RA Antoniou B., Zidanic M., Strong C., Shekher M., Matero A., Shah R.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Grant S., Shohdy N., Hasegawa A., Hameed A., Lodi M., Johnson A.,  
 RA Chen B., Maira M., Martensen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 402:769-777(1999).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=9436830; PubMed=8086440;  
 RA Horn M.A., Walker J.C.;  
 RA "Biochemical properties of the autophosphorylation of RLK5, a  
 receptor-like protein kinase from Arabidopsis thaliana.";  
 RL Biochim. Biophys. Acta 1208:65-74(1994).  
 CC -1- COFACTOR: HAVE SIGNIFICANTLY GREATER ACTIVITY IN THE PRESENCE OF  
 NM2+ THAN MG2+.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS AND ROSETTES.  
 CC -1- PTM: AUTOPOSPHORYLATED ON SERINE AND THREONINE RESIDUES.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR  
 CC FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 18 LEUCINE-RICH REPEATS (LRR).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
 CC -----  
 CC EMBL; M84660; AA32859.1; -;  
 DR EMBL; AL021749; CAAL6889.1; -;  
 DR EMBL; AL161572; CAB79651.1; -;  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF00560; LRR; 20.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 17.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 KW Transmembrane; Glycoprotein; Phosphorylation; Leucine-rich repeat;  
 KW Repeat; Signal.  
 FT SIGNAL 1 14  
 FT CHAIN 15 999  
 FT DOMAIN 15 621  
 FT TRAMEM 622 641  
 FT DOMAIN 642 999  
 FT REPEAT 88 112  
 FT REPEAT 114 138  
 FT REPEAT 139 161  
 FT REPEAT 163 186  
 FT REPEAT 188 211  
 FT REPEAT 235 261  
 FT REPEAT 263 283  
 FT REPEAT 284 306  
 FT REPEAT 307 330  
 FT REPEAT 332 353  
 FT REPEAT 354 378  
 FT REPEAT 402 426

FT REPEAT 427 450 LRR 13.  
 FT REPEAT 452 474 LRR 14.  
 FT REPEAT 498 522 LRR 15.  
 FT REPEAT 524 547 LRR 16.  
 FT REPEAT 549 567 LRR 17.  
 FT REPEAT 568 592 LRR 18.  
 FT DOMAIN 683 968  
 FT CARBOHYD 98 98  
 FT CARBOHYD 102 102  
 FT CARBOHYD 150 150  
 FT CARBOHYD 185 185  
 FT CARBOHYD 210 210  
 FT CARBOHYD 269 269  
 FT CARBOHYD 282 282  
 FT CARBOHYD 452 452  
 FT CARBOHYD 576 576  
 FT NP\_BIND 689 697  
 FT BINDING 711 711  
 FT ACT\_SITE 819 819  
 FT MUTAGEN 711 711  
 SQ SEQUENCE 999 AA; 109095 MW; F5793D89E9A0C6A7 CRC64;  
 Query Match 21.2%; Score 1311; DB 1; Length 999;  
 Best Local Similarity 32.4%; Pred. No. 2.5e-68; Indels 254; Gaps 30;  
 Matches 386; Conservative 161; Mismatches 392;  
 22 LFGQSPQSLVREIHQISFQDVDP-KNLLPMSNKN--PCTPDGVTG-RDDKYTSI 77  
 11 LSTTLPSTLSLQDATTIRQAKLSLSDPSQSLSDNDVTPCKMGLVSCDATSNVVS 70  
 78 DLSKPLNVGFSAVSSLSLTGLESFLNSHNGSVG--FKCSASLRLDLSRSL 135  
 71 DLSSFLVGFPSI---LCHLPSLHSLNYSNGSLSDPFTCHNLISLSENLV 127  
 136 GRYTTLTSLGSCGLKFLNVSNTLDPFGKXSGGLKNSLEVLDSANSIGANVGVWL 195  
 128 GSIP-----KSLPFLP----- 139  
 196 SPOGGLKHLAISGNKISGDVVSRCVNLFLDVSSNNFSGTIP-FLDCGALQHLDISG 254  
 140 -----NKLFLISGNLSDTIPESFGEFRLSLNLAG 172  
 255 NKLSDPSRAISTCTELKLNLSNQF-VGPIPLP--PLKSLQVSLANKFTGRIIPFL 311  
 173 NPLSGTIPASLGNVTLTKELKLVNLFSPSQIPSGVLTQVIMLGCNLVGPISL 232  
 312 SGACDTLTGLDSGNHFGVAVPFPFGSGSLSLASSNPFGEIPMDTLKRGQLKYL 371  
 233 S-RLTSLVMDLTFNQLTGSIPSWITQLKTVQIILFNNSFGELP-ESMGMTTLKGF 290  
 372 LSPFSGELPESLTNLSASLLTLDLSSNPFSGPILPNLCQPKNTLOELVYQNGFTGK 431  
 291 ASNKKLTGKIPDLNLT--NLSSINLFEMWLGKPLPESITR--TLTSLTKLNNRLTGV 346  
 432 IPTLSNCELYSLHSFNYLTGTPSSLSGLSKRLDKLTMNMLBGLPGLMYVYKLE 491  
 347 LPQLGANSPLQYVDLSYNRFGSEIPANVCGGKLEYLLIDNSFGEISNNLGCKSLT 406  
 492 TLIDPNDLTGELPSGLSCTNLTNMTSLNNRLTGERIPMIGLENTALIKLNNFSGN 551  
 407 RVRSLNNKLSGQIPFGFWGLPRLSLELSDNFTGSIPTTIIGAKLSLRLSKRFSS 466  
 552 IPDELGDCSLIWLIDNTMLFNGTIPAMFKOSGLIANFLAGKRYVYTKDGMKEKCHG 611  
 467 IPNEISLNGIIRISGAKNDFSGEIPESL----- 495  
 612 AGNILEFGIRSQNLRLSTRNCITTSRYGHTSPPTDNNGSMFLMSVYMLSGYIP 671  
 496 -----VKLQSLR-----LDLSKNQSLSGEIP 516  
 672 KEIGSMPLYFLNLGNHDSIGSLPDEYDRLGNIIDLSSNKLDRIPAMSNLTMVTEI 731  
 517 RELGKMLNELNLANNHLSGELPKVEGILLPVANYDLDSNQSGEIPLEQLNL-KLVNL 575

QY 732 DLNNNNLGPPEKQETPPAPFLNPPGLCGYPLPRCDPSNADGYAHQSHGRBPAS 791  
 Db 576 NLSTNHLHSGKIPLYANKTY-AHFIFNPGIC-----VLDGLCR-----K 615  
 QY 792 LAGSVAMG-----LFSFVCIPLGLLVGMRKRRKKAELMTAEGHSGSDRT 842  
 Db 616 ITRKRNIGYVILITLITLGLVGVVGIWPIAKCRRLAKKSTTLA----- 663  
 QY 843 ANNTNMLTGVEALSLNLAPEKPLKTF-----ADLLQATNGPHNSLIGSGGFGV 897  
 Db 664 ---SKM-----RSFHKLFHSHHAIADCLDEN-----VIGFGSGGV 697  
 QY 898 YKALIKDGSAAVAKKLHVSQGDRE-----FMAEMETIGIKIKARNVPLLYGCK 947  
 Db 698 YKVELRGEEVAVAKKLKSVKYGDDYSDSLNRDVFALAEVETLGTIRKHSIYRLMCCS 757  
 QY 948 VGERLLVNEVMKSGSLDYLOPKKGGVYLKLTSTRKTAIGSARGLALHNGSPHIIH 1007  
 Db 758 SGDCKLIVYEMRNGSLADYLDHGRKGGVYLGMPERLRALDAEGLSYLHHDVCPPIVH 817  
 QY 1008 RDMKSSVLLDENLEAFVSDFGMARL--MSAMDTLHSLVSTLAGTPGVPEYOSFGSGT 1065  
 Db 818 RDVSSNILLDSYGAADVAGVDFGIAXGQSGSKTPEAMGINGSCTIAPETVYTLRVNE 877  
 QY 1066 KGVYSGVVLLELNGKRTSPDGDNNLVGVV-KQHAKLRIIDVEDPELMKEDPALE 1124  
 Db 878 KSDIISFVGLVLELVGKOPTDS-ELGDKDMKWTALDKGLPEVIDPKL--DLKFK 933  
 QY 1125 IELLQHLKVAVACLDRARPRPTMVOMAFERIOGSGIDSGITRSEOG 1177  
 Db 934 EISKVIYHGLTSPPLNRPBMRKVYIMLOVSGAVPCSPNTSKSKTGG 986

## RESULT 2

CLV1 ARATH STANDARD; PRT; 980 AA.  
 ID CLV1 ARATH STANDARD; PRT; 980 AA.  
 AC Q9SY08; 004380; 0910T2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein kinase CLAVATA1 precursor (EC 2.7.1.1-).  
 GN CLV1 OR AT1G75820 OR T4012.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosida II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_Taxid=3702;  
 KY [1]  
 KY [11]  
 RS SEQUENCE FROM N.A.  
 RC STRAIN=cv. Landsberg erecta;  
 RX MEDLINE=97304386; PubMed=9160749;  
 RA Clark S.E., Williams R.W., Meyerowitz E.M.;  
 RT "The CLAVATA1 gene encodes a putative receptor kinase that controls  
 shoot and floral meristem size in Arabidopsis.",  
 RL Cell 89:575-585 (1997).  
 RN [2]  
 RS SEQUENCE FROM N.A.  
 RX MEDLINE=99178804; PubMed=10080719;  
 RA Williams R.W., Clark S.E., Meyerowitz E.M.;  
 RT "Genetic and physical characterization of a region of Arabidopsis  
 chromosome 1 containing the CLAVATA1 gene.",  
 RL Plant Mol. Biol. 39:171-176 (1999).  
 RN [3]  
 RS SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Becker U.R., Palm C.J., Federspiel N.A., Kaul S.,  
 White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 Dunn P., Egu P., Feldlyum T.V., Feng J.D., Feng B., Fujii C.Y.,  
 Gail J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,  
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 Lin X., Liu S.X., Liu Z.A., Luros J.S., Maltz R., Marzilli A.,  
 Miltsecher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.T.,  
 Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 Sakano H., Salzberg S.L., Schwartz U.R., Shin P., Southwick A.M.,  
 Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 Utterback T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,  
 Wu D., Yu G., Fraser C.M., Venter R.W.,  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 408:816-820 (2000).  
 CC -1- FUNCTION: ACTS WITH CLV3 TO CONTROL THE BALANCE BETWEEN MERISTEM  
 CELL PROLIFERATION AND DIFFERENTIATION. MAY ACT WITH CLV3 AS A  
 CC LIGAND-RECEPTOR PAIR IN A SIGNAL TRANSDUCTION PATHWAY,  
 CC COORDINATING GROWTH BETWEEN ADJACENT MERISTEMATIC REGIONS.  
 CC -1- SUBUNIT: MULTIMER (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).  
 CC -1- TISSUE SPECIFICITY: IN A CENTRAL REGION OF THE SHOOT AND IN EARLY  
 CC FLOWER MERISTEMS.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 18 LEUCINE-RICH REPEATS (LRR).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC  
 DR EMBL, U96879; AA58929.1; -  
 DR EMBL, AF049870; AAD02501.1; -  
 DR EMBL, AC007396; AA26772.1; ALT INIT.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam, PF00069; pkinase; 1.  
 DR Pfam, PF00560; LRR; 18.  
 DR PRINTS; PR00019; LRRICHRPT.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SMO0370; LRR; 18  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP, FALSE\_NEG.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST, 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Receptor; Differentiation; Signal; Repeat; Leucine-rich repeat;  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Phosphorylation; Transmembrane.  
 FT SIGNAL 1 24  
 FT CHAIN 25 980  
 FT DOMAIN 25 638  
 FT TRANSMEM 639 659  
 FT DOMAIN 660 980  
 FT REPEAT 93 119  
 FT REPEAT 143 167  
 FT REPEAT 168 193  
 FT REPEAT 240 264  
 FT REPEAT 265 288  
 FT REPEAT 289 312  
 FT REPEAT 314 336  
 FT REPEAT 337 360  
 FT REPEAT 361 384  
 FT REPEAT 386 408  
 FT REPEAT 409 432  
 FT REPEAT 435 459  
 FT REPEAT 480 503  
 FT REPEAT 505 527  
 FT REPEAT 528 551  
 FT REPEAT 552 575  
 FT LRR 1.  
 FT LRR 2.  
 FT LRR 3.  
 FT LRR 4.  
 FT LRR 5.  
 FT LRR 6.  
 FT LRR 7.  
 FT LRR 8.  
 FT LRR 9.  
 FT LRR 10.  
 FT LRR 11.  
 FT LRR 12.  
 FT LRR 13.  
 FT LRR 14.  
 FT LRR 15.  
 FT LRR 16.  
 FT POTENTIAL.  
 FT RECEPTOR PROTEIN KINASE CLAVATA1.  
 FT EXTRACELLULAR (POTENTIAL).  
 FT POTENTIAL.  
 FT CYTOPLASMIC (POTENTIAL).

FT REPEAT 577 600 LRR 17.  
 FT REPEAT 675 691 LRR 18.  
 FT DOMAIN 692 698 PROTEIN KINASE  
 FT NP BIND 698 706 ATP (BY SIMILARITY)  
 FT BINDING 720 720 ATP (BY SIMILARITY)  
 FT ACT SITE 817 817 BY SIMILARITY  
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 510 510 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CONFLICT 235 235 R -> P (IN REF. 3)  
 FT CONFLICT 703 703 A -> S (IN REF. 2)  
 FT CONFLICT 856 856 G -> D (IN REF. 2)  
 SQ SEQUENCE 980 AA; 107656 MW; B69B2DA7A8863295 CRC64;

Query Match 19.8%; Score 1225.5; DB 1; Length 980;  
 Best Local Similarity 29.4%; Pred. No. 2.1e-63;  
 Matches 352; Conservative 189; Mismatches 358; Indels 297; Gaps 35;  
 14 LFPSPFSLFQASPSQSLREIHOLISFDVLT--PDKNLIPDMSSKNP---CTPDGYLT 68  
 13 LHLVLFSPPC-----AYTDMEVLLNLKSSMIGPGHGLHMDHSHSSPDACSFSGS 65  
 69 CRDD-KTSTIDLSKPLNPGSAVSSSLSTGLSESLFLSNHSHGVSQFSGKASLTSL 127  
 66 CDDARVLSLNVSTPL--FGTISPEIGMLTHLVNLTLANNTG-----EL 110  
 128 DLSRNSLGPVTLTSLGSCGLKFLNVSNN--TLDPKGVSGGLKLNLEVLDSANS 184  
 111 PLEMSL-----TSLKVLNLSNNGNLTGTFPGEI----- 139  
 185 ISGANVAVWVSDGCGELKHLAISNKISGVDVSRVCNLEFLVSSNPFSTGIPFLGDC 244  
 140 -----LKAAMDLEVLTYNNNN----- 157  
 245 SALQHLDISGNKLSGDSFRAISTETELKLNLSNQVGPPIP--LPLKSLQVLSLAENK 302  
 158 -----GKLPEMSELKCLKYLSPGNGF 179  
 303 FTGEIPDPLSGACDTLTGLDLSGNHFGAVPPFGSCSLSLALS--SNFSGELPMD-- 359  
 180 FSGEIPESY--GDIOSELYGLNAGLSGKSPAFSLRKLRLREMYGYNSYTGAVPREFG 238  
 360 TLLKRGKLVLDLSPREPSGELPESLTLNLSASLTLDLSNNFSGPILPNLCQNPNTLQ 419  
 239 GLTK--LEILMASCTLTGELPSTLSN-----KTLH 268  
 420 EYLQNGNGTGKIPPTLSCSEIVLSHLSFNYSGLTIPSSLSKLDKLMMLMGE 479  
 269 TLFHLNNTLGHIPPLSGVLSKSLDLSINQLTGELIPQSFNLNITLINFRNNLTGQ 328  
 480 IPOELMYVTLETLTLDLDFDNLGEISGSNCTNINWLSLNNRNLGTGELPKMIGLEMLA 539  
 329 IPEALIELPGLVFEVREWENNFTLQPLANIRGNGLIKLVDSDHLTGLTPKOLCGEKLE 388  
 540 IIKLSNNSGNI PDELGDCRSILWLDLNTLNGTIPAMFK-----QSGKIANFIACK 595  
 389 MLLSNPFPGIPPEIKGCKSLTKIRYKNLNGVPGFLPLVITIELDLNDFESGE 448  
 596 RYVYIKNDGKKECHGAGMLLEFGIARSEQNLSTRNPNCTISRYVGGHTSPTPNNKS 655  
 449 LPTVMSGDVL-----DQIYLSNNFSGEIPPAIGNFBN 481

QY 656 MMFLDMSYNNLSGYIKETGSMPLYLFIINLGHNDISGSIIPD----- 697  
 DB 482 LQTLFDLRNFRFNIPREIFELHLSRITSANNITGGIPDSISCRSTLISVDSLRNN 541  
 QY 698 -----VGLDRGINIDLSNKLDRIPQAMGALTMTLEIDLSNNLSGPIPMQGFET 751  
 DB 542 GEIPKGINNVKNGLTINISGNQLTGSIPTGIGMTSLTLDLSFDLSGRVPLGQFLVF 601  
 QY 752 PPAKFLNPGLCYPLPR---CDPSNADGYAHQSHGRPPASLAGSVAMGLFSPVCTF 808  
 DB 602 NETSFANVTYLC---LPHRVSC-PTRPQOTSDHNTALFSPSRITV-----IALIT 650  
 QY 809 GLTIVGEMKRRKRKEAELEMYAEGHNSGDRTANNTWKLTVGEALSLINLAPEKEL 868  
 DB 651 GLTILSVAILQMNKKKXK-----SLAMKLT-----AFQ--- 679  
 QY 869 RKLTF--ADLLQATNGPHNDLSIGSGFGDVYKALLKDSANAIKULI-HVSGQGREPM 925  
 DB 680 -KLPFKSEDLVLEC---LKENIIGKGAGIVYRGSMPNNVDVAIKELVGRGTRSDHGT 735  
 QY 926 AEMETGKIKRNVPLTGYCKYGDDEBLVNEVWKGSLEEDVADPPKGGVKKLSTRK 985  
 DB 736 AEIQTGIRIRHATVRLGLVANKDINLLYETPNPSLGEILHSGSG--HLQWETRRH 793  
 QY 986 IAIQSARGLAFLHNCSPHIIRDMKSSNYLLDENTLEAVRVSDFGMALMSAMDTHLSVT 1045  
 DB 794 VAVRAKGLCYLHNDGCPILHRYKSNNTLLDSDFAHVADEFLAFLVDGAASECMSS 853  
 QY 1046 LAGTGVVPPRYGOSFRCTGYGYVGVVTLLELTKRPDSDPDDN-NLVGMKQHA 1104  
 DB 854 IAGSYGYIAPVAVATLKVDEKSDVYSFGVVLLELTAICKRPGV--EPESGVDYANVR-NT 910  
 QY 1105 KLRISDVDPDELMEK--DPALE-----IELLOHLKVAVACLDDBRAMRPTVQVAM 1154  
 DB 911 EELITQPSDAIIVAVIADPRLTGYPLISVHVFRIAMWCVEEAAPPTREYVHM 966  
 RESULT 3  
 ID TMKL ARATH STANDARD; FRF; 942 AA.  
 AC P43298;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative receptor protein kinase TMKL precursor (EC 2.7.1.-).  
 GN TMKL OR A01G65150 OR P15E12.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=93076110; PubMed=1332795;  
 RA Chang C., Schaller G.E., Patterson S.E., Kwok S.F.,  
 RA Meyerowitz E.M., Blecker A.B.;  
 RT "The TMKL gene from Arabidopsis codes for a protein with structural  
 and biochemical characteristics of a receptor protein kinase.";  
 RL Plant Cell 4:1263-1271 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Corn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feildlyum T.V., Feng J.-D., Feng B., Fujii C.Y.,  
 RA Gail J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marshall A.,  
 RA Millicher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pat G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tamunga G., Toriumi M.J., Town C.D.,  
 RA Utecherack T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.,  
 RT "sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:816-820(2000).  
 CC -1- FUNCTION: PROBABLE RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.  
 CC -1- PTM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 9 LEUCINE-RICH REPEATS (LRR).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: L00670; AAA32876.1; -  
 DR EMBL: AC026480; AAG51302.1; -  
 DR HSSP: P12931; IFMK.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR InterPro: IPR003591; LRR\_typ.  
 DR InterPro: IPR004040; STY\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00560; LRR; 11.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00369; LRR\_Typ; 1.  
 DR SMART: SM00221; STYKc; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR Transfaser: Serine/threonine-protein kinase; ATP-binding;  
 KW Transmembrane; Receptor; Glycoprotein; Signal; Repeat;  
 KW Leucine-rich repeat; Phosphorylation.  
 FT SIGNAL 1 23  
 FT CHAIN 24 942  
 FT DOMAIN 24 482  
 FT TRANSMEM 483 503  
 FT DOMAIN 504 942  
 FT REPEAT 87 110  
 FT REPEAT 111 133  
 FT REPEAT 135 159  
 FT REPEAT 185 209  
 FT REPEAT 231 253  
 FT REPEAT 254 278  
 FT REPEAT 280 300  
 FT REPEAT 385 408  
 FT REPEAT 409 436  
 FT DOMAIN 588 869  
 FT NP\_BIND 594 602  
 FT BINDING 616 616  
 FT ACT\_SITE 717 717  
 FT CARBOHYD 86 86  
 FT CARBOHYD 99 99  
 FT CARBOHYD 158 158  
 FT CARBOHYD 164 164  
 FT CARBOHYD 171 171  
 FT CARBOHYD 230 230  
 FT CARBOHYD 285 285  
 FT CARBOHYD 363 363

FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 942 AA; 102387 MW; 93E300B52FF5499E CRC64;  
 Query Match 13.5%; Score 834; DB 1; Length 942;  
 Best Local Similarity 26.3%; Pred. No. 9,1e-41;  
 Matches 293; Conservative 177; Mismatches 373; Indels 272; Gaps 38;  
 QY 153 LANSSTLPPPGVSGGLKLNLEVLDSANGISGANVGVVLSGCGELHGLAISGNKI 212  
 DB 16 LILSLKADSDGLSAMLSLK-----KSLNPPSFFGMSDPPC-KWHIYCTGTR- 64  
 QY 213 SGDVVSRCVNLEFLVSSNNFSTGIPFLGDCSALQHDIDGNKLSGDFSAITGTELK 272  
 DB 65 -----RTRQIGHSGLQTLSPDLNLSBL 91  
 QY 273 LNISSNQVGPPLP-LKSLQYLSLANKFTGTEIPDLGACDTLTGLDGNHFG- 330  
 DB 92 RLELQNNISGVPFSLGSLAQVLMLSNNF-DSIPEDVFOGLTSLQSVRIDNNPKSW 150  
 QY 331 AVPPFGCSLSESLALSNNFSGELP---MDTLKRGKXVLDLSPNERSGLPESLT 386  
 DB 151 EIPESLRNMSALQNSANSANVSGSLPGLPDD---EPGSLIHLAPNMLEGSLPMSLA 207  
 QY 387 NLSASLITLIDSSNNFSGPLPNIQNPKNLTQELYLQNNGFTGKIPPLTSCSELSLH 446  
 DB 208 G-----SQVSLMLNGQLTGDITVQNTGKLEW 238  
 QY 447 LSPNYLSGTPISGLSKLADKLMLNMLEGELPQELMAYKTLTETLIDPNDITGEIPS 506  
 DB 239 LSNKRFSGPLPPSG-----LKELESLSLRNDSFTGVPVA 273  
 QY 507 GLSNCNLNWLISLNNRLTGEIPKMGLEVLALIKLSNNSFGSNIPELDGDC---RSL 562  
 DB 274 SLTSLSESLKVNLTNNHLQGPVYFKSSV--VLDKDSNFCLSG--GECDPYKSL 328  
 QY 563 TWDLNTLNFNGTIPAMFKOSGKIAPNFIAGKRYVYIKDKMKKECHGAGNLEFGAIR 622  
 DB 329 LLI-----ASSFDYPPRLAESW-----KGN----- 348  
 QY 623 SEQLNRSLTRNPNITSNVYGHSTPFDNNGSMFLDMSYMLSGTIPETISMPYLF 682  
 DB 349 -----DPCF-----NMGISAGNNTVYSLKMELTGTISPEFAIKSLOR 390  
 QY 683 LNLGNDISGIPDEVGLRNLNIDLSNKLDRIPQASALTMLTLEIDLSNNISGLP 742  
 DB 391 IILGNNLTGMIPQELTLPVLKTLVDVSNKLFGKVPGRFSNVVNT-----NGN----- 440  
 QY 743 PEMQGFETFPFAKFLNPNGLCGYLPFCDDPSNADGYAHHQRSHGRPAS-----LAGSV 796  
 DB 441 PDIGDKS-----SLSPG-----SSSPSGSGSGINGDKDRGMKSTFGLIVGSV 488  
 QY 797 AMGLLFSPVCFGLILVGRMRKRKRKEALEMYAEGH-GNSGDPTANTNTMKTGYKE 855  
 DB 489 LGLLSLIF--LIGLVFCWYKKQKRFSGSSSNNAVVPHRSG--SNESEKTVAGS 543  
 QY 856 ALST-----NLAEPEKPLRKLTFADLLQATNGFNDLSISGGFGVY 898  
 DB 544 SVSYGSGIDFTYLLPTSEVGDNIQWEAGNMLISIQVLSVNNFSSDILISGGFGVY 603  
 QY 899 KAILKDSAVAIKKLTH--VSQGDREFAEMETTKIKIHRNLVPLLGYCKYGDRELLVN 956  
 DB 604 KGLHDSGTAKVAKRMKMGVLAGGFAEFSEIAVLTKVHRHLVTLVGLYCLDGNKELVY 663  
 QY 957 EVMKYSGLIEDVLDPPKKGVK-LKLSRKRKLAGSARGLAFLHNHNSPHI IHRDKSSNV 1015  
 DB 664 EYMPQGLTSHLFEWSEBGLKPLMKQRLTALDVARGVYIHLGALHGSFHRDLKPSNI 723  
 QY 1016 LLDNLEEARSDGMARLMSANDTHLSVST-LAGTPGVYPPPYGSPFCSTKGDVYSYGV 1074  
 DB 724 LLDGDMAKVADGLVRL--APEGKSIETRIAGTFGYLAPEYAVTGVTTKVDVYSYGV 781  
 QY 1075 VLEELTGRKPTSPDFGNN-NLVGVVVK-----HAKLRSDVFDLEMKEDPALTE 1126

Db 782 LMELITGRKSLDESDQPEESHLVSWFKMYINKKASFKKADITDID---LDEETLASVH 838  
 Qy 1127 LLOHLKVAVACDDRAMRRPTM---VQVAMFKEIQAGS-----GID----- 1165  
 Db 839 TVAEI--AGHCARREPQRPDMGHAIVNLSLVLEWKPSPQNPEDYIGDIDLMSPQALK 896  
 Qy 1166 -----SQSTIRSEIDGGFSTIEMVMSIKEVEEG 1194  
 Db 897 KMQAVEGRSDLESSTSLPSLDNTQMSIPTRPYG 931

RESULT 4  
 ID TML1 ARATH STANDARD; PRT; 674 AA.  
 AC P33543;  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 41, Last annotation update)  
 DE Putative kinase-like protein TML1 precursor.  
 DE TML1 OR AT3G24660 OR MSD24.3.  
 OC Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE=Green siliques;  
 RX MEDLINE=9403320; PubMed=8219075;  
 RA Valon C., Smalle J., Goodman H.M., Giraudet J.,  
 RT "Characterization of an Arabidopsis thaliana gene (TML1) encoding a  
 RT putative transmembrane protein with an unusual kinase-like domain.",  
 RL Plant Mol. Biol. 23:415-421 (1993).  
 RT [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=2036309; PubMed=10907853;  
 RA Kaneo T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the 4,251,695 bp regions covered by 90 pl. TAC  
 RT and BAC clones".  
 RL DNA Res. 7:217-221 (2000).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: TO PROTEIN KINASES IN THE C-TERMINUS. BUT DOES NOT  
 CC SEEM TO HAVE CONSERVED A KINASE ACTIVITY.  
 CC -1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC  
 DR EMBL: X72863; CAAS1385.1; -  
 DR EMBL: AP000740; BAB01215.1; -  
 DR PIR: S35397; S35397.  
 DR PIR: S39476; S39476.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00560; LRR; 5.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00370; LRR; 4.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 DR Receptor; Transmembrane; Glycoprotein; Signal; Leucine-rich repeat;  
 KW Repeat.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 674 PUTATIVE KINASE-LIKE PROTEIN TML1.  
 FT DOMAIN 26 295 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 296 323 POTENTIAL.

FT DOMAIN 324 674 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 373 674 PROTEIN KINASE LIKE.  
 FT REPEAT 98 122 LRR 1.  
 FT REPEAT 123 146 LRR 2.  
 FT REPEAT 147 170 LRR 3.  
 FT REPEAT 172 195 LRR 4.  
 FT REPEAT 198 224 LRR 5.  
 FT REPEAT 226 245 LRR 6.  
 FT REPEAT 246 268 LRR 7.  
 FT CARBOHYD 57 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 90 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 674 AA; 73353 MW; 5330DA8D8767E253 CRC64;

Query Match 10.3%; Score 639; DB 1; Length 674;

Best Local Similarity 27.6%; Pred. No. 1,1e-29; Mismatches 254; Indels 156; Gaps 25;

Matches 209; Conservative 137; Mismatches 254; Indels 156; Gaps 25;

Qy 445 LHLSPNYL-----SGTIPSSIGSLKRDIAKMLNMEGIP--QELMYKTEITLI-- 494  
 Db 9 LHVIFPVLILHCHCGT-----SLSGSDVKKLLIGKIKSLQGNSSBLSLWSNNSVVP 62  
 Qy 495 -----LDNDLTGE--IPSGLSNCPNINWISLS--NNRLTGEIPKWIQRL 535  
 Db 63 COMRGKWFVNSGSPLOQSDSSPQWNTSLFNDSHLHLSQLPSANLTSLSREIGEF 122  
 Qy 536 EKLATIKLSNNSPFGNIPDELCDCLRLIWLDTNLTPNGTIPAAFKQSGKIAANFLAGK 595  
 Db 123 SMLQSVFLNINLSGSIPLHELGYTSSLDVDSGKALAGVLPSPIMNCDKL----- 174  
 Qy 596 RYVYIKNDMKKECHGAGNLLEFGIRSEQLNRLSTRPNCITSRVYVGHSPPTDNGS 655  
 Db 175 -----VSFKIHG----- 181  
 Qy 656 MNPFLMSYMLSGYIPKEIGSWP-----YLFILNCHNDISGIPDEYDGLNLIIDLS 710  
 Db 182 -----NNLSGLVLEP--ALPNSTCGNLQVLDGSKNKFSEFEPFETTRFGVAGSLDS 231  
 Qy 711 SNKLDRIPOAMSAITMLTEIDLNNNLSGPIPEMGQETPPPAKFLNPPGLCGYPLPRC 770  
 Db 232 SNVFBELVVEGLGVLE--LESLNLSHNFSGMLPDGSEKFGAEGSGNSPSLCGLPLKPC 290  
 Qy 771 DPSNADGVAHHQSHGRPRASLAGSVAMGLFSPYCIFELLVGEREMRRKKEALEM 830  
 Db 291 LGSRR-----LSPGAVAGLV--IGLMSGAVVAVASLLGYLQNKRRKSSISEDDL 338  
 Qy 831 YAEHGNSGDRNTANNNTNMTLGVKEALSINLAFEKPLKLTFFADLLQATNGFHNDSLIG 890  
 Db 339 -----EEGDEE-----IGKEGGBEKLIVFQGG--ENLTLDDVLANTG-----QVME 380  
 Qy 891 SGGFQDVYKAILKDGSAVAIKKLHVSGQDREPMAMETIGIKIRNVLPLGIC--KV 948  
 Db 381 KTSYGTVYKAKLSDGNNILRLIREGTCXDRSSCLPIVIRGLIRHENLVPRLAFQGR 440  
 Qy 949 GDEBLVNEVMKSGSLIEDVLOPKKGVKLKSTRKIKIGSARGAFPHNCSPIHRR 1008  
 Db 441 G-EKLLIYDLPNISLHDLHESKPRKPPALNMWRKIKMLGILKGLAYLHTQGEVITIG 499  
 Qy 1009 DMKSSNVLDLENLEARVSDGMARLM--SAMDTHLSVSTLAGTPGVVPEYQSFRCSTK 1066  
 Db 500 NIKSKNVLDVDFPARLTERGDKIMVQAVAD--EIVSQAKSDGKAKDELHMKKCNRR 556  
 Qy 1067 GDVYSYGVLEILLTGRRPDSPPDGN--NLVGYWKQHA--KLRISDVDPPELMK--EDP 1121  
 Db 557 SDVYAFGLILELILMGKPKSGKNGNEFVDLPSTLYKAALTEETWEVDLEAMKGRSD 616  
 Qy 1122 ALLEILLKLVAVACDDRAMRRPTMVQVAMFKE 1157



InterPro; IPR004040; STY\_pkinase.

DIFFERENTIATION SIGNAL. THE CRINKLY4 (CP4) MUTATION AFFECTS THE FUNCTION: POSITIVE RECEPTOR PROTEIN KINASE. COULD PLAY A ROLE IN A

EPIDERMIS DIFFERENTIATION SUCH THAT CELL SIZE AND MORPHOLOGY ARE

ALTERED, AND SURFACE FUNCTIONS ARE COMPROMISED, ALLOWING GRAFT-LIKE FUSIONS BETWEEN ORGANS.

CC - SUBCELLULAR LOCATION: Type 1 membrane protein.

CC - SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC - SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.

CC - This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U67422; AAB09771.1; -

DR MafzDB; 128723; -

DR InterPro; IPR000719; Euk\_pkinase.

DR InterPro; IPR004040; STY\_pkinase.

DR InterPro; IPR002290; Ser\_thr\_pkinase.

DR InterPro; IPR001368; TNFR\_c6.

DR Pfam; PF000020; TNFR\_c6; 1.

DR Pfam; PF000069; pkinase; 1.

DR ProDom; PD000001; Euk\_pkinase; 1.

DR SMART; SM00221; STYKc; 1.

DR SMART; SM00208; TNFR; 1.

DR PROSITE; PS00107; PROTEIN KINASE\_ATP; 1.

DR PROSITE; PS00108; PROTEIN KINASE\_ST; 1.

DR PROSITE; PS50011; PROTEIN KINASE\_DOM; 1.

DR PROSITE; PS00652; TNFR\_NGFR\_1; FALSE\_NEG.

DR PROSITE; PS50050; TNFR\_NGFR\_2; 1.

DR Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat; Transmembrane; Signal.

KW SIGNAL

FT CHAIN 1 24

FT CHAIN 25 901

FT TRANSF 25 423

FT TRANSF 424 444

FT TRANSF 445 901

FT TRANSF 33 330

FT TRANSF 33 68

FT TRANSF 72 107

FT TRANSF 125 160

FT TRANSF 162 195

FT TRANSF 203 236

FT TRANSF 253 287

FT TRANSF 292 330

FT TRANSF 357 391

FT TRANSF 505 511

FT TRANSF 511 519

FT TRANSF 533 533

FT TRANSF 634 634

FT ACT SITE 901 AA; 97439 MW; 5218481AC187E061 CRC64;

SEQUENCE

Query Match 8.5%; Score 528; DB 1; Length 901;

Best Local Similarity 39.8%; Pred. No. 4.3e-23;

Matches 131; Conservative 63; Mismatches 101; Indels 34; Gaps 13;

QY 841 RTAANT-----NMKLTGVKALSTINLAFAEKPLKLFADLLQATNGFNHDSLGSG 893

DB 462 RLAKSTAVSEFRKDKMKTQPMEDLKTIRA-----QESFSEIEQKLTGFSSEDSOVGKGS 515

QY 894 FGQVYFAILKDGSAVAIKLIHVS--QGDREPMAMETIGIKRNLVPLLGYCKVGD 951

DB 516 FSCVFGKILRDGTVAVAKIKASDVKSKSEPHNLDLSRLNHAHLNLTGCEGDS 575

QY 952 RLIVNEVMKYSLEVL--QDP--KKGAVLKLIKSTKRLAIGARGLATLHNCSHITH 1007

DB 576 RLIVNEVMKYSLEVL--QDP--KKGAVLKLIKSTKRLAIGARGLATLHNCSHITH 1007

QY 1008 RDKKSSNVLLDEMLARVSPGEMARLMSAMDTHLSVTL--AGTPGVVPEYVYQSFRCSTK 1066

DB 633 RDKKSSNVLLDEMLARVSPGEMARLMSAMDTHLSVTL--AGTPGVVPEYVYQSFRCSTK 1066

QY 1067 GDVYSVGVLELITLTKRPPDPSDPFGDNMLVGWYKQAKRLISDFV---DPELMKEDPAL 1123

DB 692 SDVYSFVLELITLTKRPPDPSDPFGDNMLVGWYKQAKRLISDFV---DPELMKEDPAL 1123

QY 1124 EIELLOHL-KVAACLDLDRMRPTWQV 1151

DB 746 DLEALKKIASVACVCKVEMRGKDRPSMDKV 774

RESULT 7

APKA ARATH STANDARD; PRT; 410 AA.

ID APKA ARATH

AC Q06548;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protein Kinase APKA (EC 2.7.1.1-).

GN APKA.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Eusteroideae; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI\_Taxid=3702;

OK NCBI\_Taxid=3702;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=93081726; PubMed=1450380;

RA Hirayama T., Oka A.;

RT "Novel protein kinase of Arabidopsis thaliana (APKA) that phosphorylates tyrosine, serine and threonine."

RL Plant Mol. Biol. 20:653-662(1992).

CC - FUNCTION: POSSIBLE BI-FUNCTIONAL KINASE. IN VITRO, IT EXHIBITS SERINE/THREONINE ACTIVITY. IN VIVO, CAN PHOSPHORYLATE TYROSINE RESIDUES OF LIMITED SUBSTRATES.

CC - SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; D12522; BAA02092.1; -

CC InterPro; IPR000719; Euk\_pkinase.

CC InterPro; IPR004040; STY\_pkinase.

CC InterPro; IPR002290; Ser\_thr\_pkinase.

CC Pfam; PF000069; pkinase; 1.

CC ProDom; PD000001; Euk\_pkinase; 1.

CC SMART; SM00221; STYKc; 1.

CC PROSITE; PS00107; PROTEIN KINASE\_ATP; 1.

CC PROSITE; PS00108; PROTEIN KINASE\_ST; 1.

CC PROSITE; PS50011; PROTEIN KINASE\_DOM; 1.

CC Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase; ATP-binding; Multigene family; Myristate.

KW ATP-binding; Multigene family; Myristate.

FT DOMAIN 2

FT DOMAIN 68 352

FT NP BIND 74 82

FT BINDING 106 106

FT ACT SITE 203 203

FT ACT SITE 410 AA; 45519 MW; 5BAB28D9E0065082 CRC64;

SEQUENCE

Query Match 8.2%; Score 508; DB 1; Length 410;

Best Local Similarity 36.5%; Pred. No. 2.1e-22;

Matches 129; Conservative 61; Mismatches 131; Indels 32; Gaps 9;

QY 868 LRKLPADLLQATNGFNHDSLGSGGVDYKALIND-----GSAVAIKLIHVS 917

DB 53 LKSFSEALSKATNFPDPSDPFGDNMLVGWYKQAKRLISDFV---DPELMKEDPAL 1123

QY 918 GQGDREPMAMETIGIKRNLVPLLGYCKVGDRLIVNEVMKYSLEVLQDPKKGAVK 977

Db 113 WQHQEWLAENVYLGPSHRLVYLICLEDERHLLVYFMRPGSLIENHLLFRGGYFOP 172  
 Qy 978 LKSTRRKIAIGSARGAFLHNCSPHIIHRDMKSNVLLDENLEAVSPDGMARLMSAM 1037  
 Db 173 LSWRLRLKVALGAAGKGLAFL-HSEETRYIYRDFKTSNILLDSYNAKLSDFGLAKGPG 231  
 Qy 1038 D-THLSVSTLAGEGYVPEYVYOSFRSTGKDYVSVVTLLELTGKRPD-SPDFGDN 1095  
 Db 232 DKSHVSTRVM-GTIGYVAPPEYLAATGLTTSVDVSVFVGLLELISGRRAVDKRPSEGRN 290  
 Qy 1096 LVGVKQHA-KRISDPFPELMKEDPALEIELQHLKVAACLDRAVRPPTVQVNA 1153  
 Db 291 LVEMAKRYLVKRIKFRVIDNRL--QOYSMERACVATLSRLCTTEIKAPNMSSEVS 348  
 Qy 1154 MFKEIOA-----SGSID-----SQTIRSEIDGGFSTTEMVMSIKEVP 1192  
 Db 349 HLEHTIQSLNAIGNMKIDTRMRRRSDSVSKVNAAGPRAVAGSTVAVP 401

RA 8  
 RA 8 ARATH  
 RA 8 APKB ARATH STANDARD; PRT; 412 AA.  
 AC P46573; Q9SLH5;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein kinase APK1B (EC 2.7.1.-).  
 GN APK1B OR AT2G28930 OR T914.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxId=3702;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Coppenhaver G.P., Preuss D.,  
 RA Nieman W.C., White O., Eisen J.A., Salzberg S.L., Frazer C.M.,  
 RA Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RA thaliana.";  
 RT Nature 402:761-768(1999).  
 RA [12]  
 RC SEQUENCE OF 143-346 FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA MEDLINE=93081726; PubMed=1450380;  
 RA Hirayama T., Oka A.;  
 RA "Novel protein kinase of Arabidopsis thaliana (APK1) that  
 RT phosphorylates tyrosine, serine and threonine.";  
 RL Plant Mol. Biol. 20:653-662(1992).  
 CC -1- FUNCTION: POSSIBLE BI-FUNCTIONAL KINASE. IN VITRO, IT EXHIBITS  
 CC SERINE/THREONINE ACTIVITY. IN VIVO, CAN PHOSPHORYLATE TYROSINE  
 CC RESIDUES OF LIMITED SUBSTRATES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AC005315; AAC33221.1; -  
 DR EMBL; D10152; BAA20968.1; -  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR004040; STY\_kinase.

DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00669; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 KW transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;  
 KW ATP-binding; Multigene family; Myristate.  
 FT IPRID 2 2 MYRISTATE (BY SIMILARITY).  
 FT DOMAIN 69 356 PROTEIN KINASE.  
 FT NP\_BIND 75 83 ATP (BY SIMILARITY).  
 FT BINDING 107 107 ATP (BY SIMILARITY).  
 FT ACT\_SITE 204 204 BY SIMILARITY.  
 SQ SEQUENCE 412 AA; 45746 MW; EBICA0B1A626A5DA CRC64;

Query Match 8.2%; Score 507.5; DB 1; Length 412;  
 Best Local Similarity 36.2%; Pred. No. 2.2e-22;  
 Matches 138; Conservative 68; Mismatches 126; Indels 49; Gaps 15;

Qy 830 MTAEGHNSGDRTPAN--NTMVKLTGVKEAL-SINLAFFKPLRKTFFADILQATNGFEHN 885  
 Db 21 MSESANDSISGKSSVSIRTPRTG--EILQSPNLSF-----TFALKATNFRP 71  
 Qy 886 DSLISGGGGRDYKALIND-----GSAVAIKLIHVSOGGDEPFMAENETGKIR 935  
 Db 72 DSVLGGGGGVSFKKMIDEQTLTKSPGTGVIAVKTLMDQGWQHQEWLAENVYLGQFS 131  
 Qy 936 HNNVPLLYGCVKGERLILNENKYGSLBDVLDOPKKGV--KLKSTRKRLAIGARG 993  
 Db 132 HNNVPLLYGCVKGERLILNENKYGSLBDVLDOPKKGV--KLKSTRKRLAIGARG 189  
 Qy 994 IAFIHHNCSPHIHRDMKSNVLLDENLEAVSPDGMARLMSAM--THLSVSTLAGEPY 1052  
 Db 190 IAFI-HNSETSYIYRDFKTSNILLDSYNAKLSDFGLAKGPGDKSHVS--TIMGTGYG 247  
 Qy 1053 VPEYVYOSFRSTGKDYVSVVTLLELTGKRPD--PPFGDNILVGVKQ--HAKLRI 1108  
 Db 248 AARPEYLAATGLTTSVDVSVFVGLLELISGRRAVDKRP--GQKILVEMARPLANKRL 306  
 Qy 1109 SDVPPELMKEDPALEIELQHLKVAACLDRAVRPPTVQVNAFKEIOAGSGIDSGS 1168  
 Db 307 FRVIDNRL--QOYSMERACVATLSRLCTFEIKAPNMSSEVSHLEHQ----- 355  
 Qy 1169 TIRSEIDGGFSTTEMVMSIKEVP 1192  
 Db 356 ---TLNAGRNNDVQRRMR 373

RESULT 9  
 XPRO\_MAIZE STANDARD; PRT; 817 AA.  
 ID XPRO\_MAIZE  
 AC P17801;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative receptor protein kinase ZMPK1 precursor (EC 2.7.1.37).  
 GN PK1.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC NCBI clade; Andropogoneae; Zea.  
 OX NCBI\_TaxId=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. B73; TISSUE=Root;  
 RA MEDLINE=90294911; PubMed=2163028;  
 RA Walker J.C., Zhang R.;  
 RA "Relationship of a putative receptor protein kinase from maize to the  
 RT S-locus glycoproteins of Brassica.";  
 RL Nature 345:743-746(1990).

[2] SEQUENCE FROM N.A.  
 RC STRAIN=CV. B73;  
 RA Zhang R., Walker J.C.,  
 RL Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLE RECEPTOR. INTERACTION WITH A LIGAND IN THE  
 CC EXTRACELLULAR DOMAIN TRIGGERS THE PROTEIN KINASE ACTIVITY OF THE  
 CC CYTOPLASMIC DOMAIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. THE SHOOTS AND  
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE ROOTS OF THE  
 CC ROOTS OF YOUNG MAIZE SEEDLINGS, AND TO A LESSER EXTENT IN THE  
 CC SILKS.  
 CC -1- SIMILARITY: THE EXTRACELLULAR DOMAIN IS SIMILAR TO S-LOCUS  
 CC GLYCOPROTEINS OF BRASSICA, WHILE THE INTRACELLULAR DOMAIN IS  
 CC A SER/THR-PROTEIN KINASE CLOSELY RELATED TO RAF KINASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to a license@isb-sib.ch).

DR EMBL: X52384; CAA36611.1; -  
 DR EMBL: X67733; CAA47962.1; -  
 DR MaizeDB: 65910; IPR001480; B lectin.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR003014; PAN.  
 DR InterPro: IPR003609; Pan app.  
 DR InterPro: IPR002290; Ser thr kinase.  
 DR InterPro: IPR000858; Sloacus\_glycop.  
 DR Pfam: PF00024; PAN; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00954; S locus\_glycop; 1.  
 DR Pfam: PF01453; Agglutinin; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00108; B lectin; 1.  
 DR SMART: SM00181; EGF; 1.  
 DR SMART: SM00473; PAN\_Ap; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE: PS00111; PROTEIN KINASE DOM; 1.  
 DR Transmembrane; Serine/threonine-protein kinase; ATP-binding;  
 DR Transmembrane; Receptor; Glycoprotein; Signal.

FT SIGNAL 1 28  
 FT CHAIN 1 28  
 FT DOMAIN 29 472  
 FT TRANSMEM 473 498  
 FT DOMAIN 499 817  
 FT DOMAIN 514 817  
 FT NP\_BIND 540 548  
 FT BINDING 562 562  
 FT ACT\_SITE 658 658  
 FT CARBOHYD 83 83  
 FT CARBOHYD 128 128  
 FT CARBOHYD 228 228  
 FT CARBOHYD 279 279  
 FT CARBOHYD 329 329  
 FT CARBOHYD 339 339  
 FT CARBOHYD 452 452  
 FT CARBOHYD 91120 NM; F164B4471922E67 CRC64;  
 SQ SEQUENCE 817 AA; 91120 NM; F164B4471922E67 CRC64;

Query Match  
 Best Local Similarity 8.2%; Score 507; DB 1; Length 817;  
 Matches 187; Conservative 113; Mismatches 236; Indels 164; Gaps 30;

QY 531 WIGRLNALTLKLSNFSGNIDELDCRSLIWLIDNLTNLFN-----GITPAMAFKQS 584  
 DB 188 YIRFSDLSVLSLIV-----HVP-QVSD--IYWPDPDQNLVODGRNQYNSTRGLMTDS 238

QY 585 GKIA-ANFLAGKRYIYIK-NDGMKKE---CHAGNLLFEQGIKRE---OLNRLSTENPCNI 637  
 DB 239 GVLASDSDFDGQALVADVGPKRRLTLDPGNLRILYSNMDSDSWSVAMQPCNI 298  
 QY 638 TSHVYG-----GHTSPFEDNNGSMFELDMSYNLSGYPREISMPFLNLGNHDI 690  
 DB 299 ----HLCGPNNGI CHSPF-----PFCSCPQYATRNPGNW-----TEGCMAI 337  
 QY 691 SGSIPEVEGDLRLNLTLDLSNKLDRIRPQMSALTMFL--RIDLSNNLSGPIDEMGQF 748  
 DB 338 VNTTCRY-DKRSMRFRVRLPNTDFWGSDDQHLVSILTRCRDICSDCKCFQYQEGG 396  
 QY 749 ETRPPAKFLNPPGLCGP-----LPRCD-----PSNADGYAHQ 782  
 DB 397 SCYPAKALFSGR---TYPSDVRTIYKLPLTGVSVSNALIPRSDVDVPRRLD--CDRM 451  
 QY 783 RSHGRAP-----ASLAGSVAMGLFSPFCIFGLIIVG-----REMRRRKKKA 826  
 DB 452 NKSIRFPDPVHKTKGGESKWFYFGIAFFVVEVSFISFAWFVYLKELRPS----- 505  
 QY 827 ELEMYAGHNSGDRITANNNTWKLTGVKEALSINLAFEKPLKLTFFADLTQATNGFND 886  
 DB 506 --ELMASEKG-----YKMTSN-----FRIGYRELVAKTRPKVE 539  
 QY 887 SLISGGFGDYKAILDKGSAVAIKKLIHVSQGGDRFEMMETTGIKIRNLVPLIGYC 946  
 DB 540 --LGRGSGYIVYKVLDDHDHVAVKLENV--QKVEVQALSVIGRINMNLVRIWFGC 596  
 QY 947 KVGDERLLVNEWYKSLIEDVLDQPKKGIVYKLTSTRKLAISGARGLAFLHNHSPHI 1006  
 DB 597 SEGSHRLVSEYVENGSJANTLIFS--EGGNLTLDWEGFNIALGVAKIALYHHECLEWV 655  
 QY 1007 HRDKSSNVLLDENLEAVSDPGMARLMSAMDTHLSYSLAGTGPVYPPEYQSFRCSTK 1066  
 DB 656 HDVAPENILLDQAFEPKITPFLGLVLLNRRGSTQNVSHRGTGLYLAPEVSSLPITAK 715  
 QY 1067 GDVYSYGVLLLELTGKRPTSDPGDNNLVGWYQ-HAKLR-----I 1108  
 DB 716 VVYSYGVLLLELTGTRVSE-----LVGQDEVHSMRLKVRMLSAKLBSBEGSWI 767  
 QY 1109 SVVPELMKEPDAIEILLOKLVAVACLDPRAMRRPTM 1148  
 DB 768 DGYLDSKLR--PVNYQARTLILKLVASCLEBDSKRTM 805

RESULT 10  
 SRK6\_BRAOL STANDARD; PRT; 849 AA.  
 AC 009092;  
 DT 01-FEB-1995 (rel. 31, Created)  
 DT 01-FEB-1995 (rel. 31, Last sequence update)  
 DT 01-NOV-1995 (rel. 32, Last annotation update)  
 DE Putative serine/threonine kinase receptor precursor (EC 2.7.1.37)  
 DE (S-receptor kinase) (SRK).  
 GN SRK6.  
 OS Brassica oleracea (Cauliflower).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxId=3712;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. 5656; TISSUE=Stigma;  
 RX MEDLINE=92020942; PubMed=1681543;  
 RA Stein J.C., Howlett B., Boyes D.C., Nasrallah M.E.;  
 RT "Molecular cloning of a putative receptor protein kinase gene encoded  
 RT at the self-incompatibility locus of Brassica oleracea.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:8816-8820(1991).  
 CC -1- FUNCTION: INVOLVED IN SPOROPHYTIC SELF-INCOMPATIBILITY SYSTEM  
 CC (THE INABILITY OF FLOWERING PLANTS TO COMBINATION WITH S-LOCUS-  
 CC FERTILIZATION), PROBABLY ACTING IN COMBINATION WITH A LIGAND IN THE  
 CC SPECIFIC GLYCOPROTEINS. INTERACTION WITH A LIGAND IN THE

CC EXTRACELLULAR DOMAIN TRIGGERS THE PROTEIN KINASE ACTIVITY OF THE  
CC CYTOPLASMIC DOMAIN.  
CC - CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
CC - SUBCELLULAR LOCATION: Type I membrane protein.  
CC - TISSUE SPECIFICITY: PREDOMINANTLY IN THE PISTIL AND ANTHER.  
CC - POLYMERISM: THERE ARE A NUMBER OF DIFFERENT S ALLELES IN  
CC B. OLERACEA, POSSIBLY PROVIDING THE RECOGNITION SPECIFICITY.  
CC - SIMILARITY: THE EXTRACELLULAR DOMAIN IS SIMILAR TO S-LOCUS  
CC GLYCOPROTEIN OF BRASSICA, WHILE THE INTRACELLULAR DOMAIN IS  
CC A SER/THR-PROTEIN KINASE RELATED TO RAF KINASES.  
CC  
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION  
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION  
CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS  
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY  
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE FOR COMMERCIAL  
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M76647; AAA33000.1; ALT\_TERM.  
DR InterPro: IPR001480; B\_lectin.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR003609; Pan\_app.  
DR InterPro: IPR004040; Sty\_pkinase.  
DR InterPro: IPR002290; Ser\_thr\_pkinase.  
DR InterPro: IPR000858; Slocus\_glycop.  
DR Pfam: PF00069; Pkinase\_1.  
DR Pfam: PF00954; S\_locus\_glycop\_1.  
DR Pfam: PF01453; Agglutinin\_1.  
DR ProDom: PD000001; Euk\_pkinase\_1.  
DR SMART: SM00473; B\_lectin\_1.  
DR SMART: SM00421; STYK\_1.  
DR SMART: SM00107; PROTEIN KINASE ATP\_1.  
DR PROSITE: PS00108; PROTEIN KINASE ST\_1.  
DR PROSITE: PS50011; PROTEIN KINASE DOM\_1.  
DR Transmembrane; Serine/threonine-protein kinase; Signal; ATP-binding;  
KW Transmembrane; Receptor; Glycoprotein; Self-incompatibility.  
FT SIGNAL 1 32  
FT CHAIN 33 849  
FT DOMAIN 33 446  
FT TRANSMEM 447 466  
FT DOMAIN 467 849  
FT DOMAIN 528 779  
FT NP\_BIND 534 542  
FT BINDING 556 556  
FT ACT\_SITE 653 653  
FT CARBOHYD 47 47  
FT CARBOHYD 120 120  
FT CARBOHYD 196 196  
FT CARBOHYD 260 260  
FT CARBOHYD 314 314  
FT CARBOHYD 389 389  
FT CARBOHYD 442 442  
SQ SEQUENCE 849 AA; 97231 MW; 7E156059EDDF4370 CRC64;

Query Match 8.1%; Score 498; DB 1; Length 849;  
Best Local Similarity 23.3%; Pred. No. 2.2e-21;  
Matches 250; Conservative 145; Mismatches 336; Indels 340; Gaps 42;

QY 133 SLSGPVTTLTSGSGGLKFLVNSN-TLDPGRKY-SGGLKNSLFLVLDLSANSISGAN 189  
DB 28 ALSTIYINTLSTSTES-----LTISNKTIVSGSIFEGFRTSRWYLGMWYKVS DRT 81  
QY 190 VVGWV-----LSDGCGELKHLAISGNKISGDVDVSRVNLLEFLDVSSNNFSTGIPPLGD 243  
DB 82 YV-WVANDNPLSNALITLK--ISGN-----NLVLLDHSN----- 113  
QY 244 CSALGHLISGNKLSGDFSRALSTCTELKLNISNNOFVGIPLPLPKSLQVYSLAE-NK 302  
DB 114 -----KPVWWTNLTNRGNE 126

QY 303 FTGEIPDLISGACDTLGLTGLDLSGNHFYGAVPFPGSCSLLESIALSSNPSGE----- 355  
DB 127 RSPVAEHLA-----NGNFV-----WRDSNNDAASYLMQSD 159  
QY 356 LPMDFLLMRGLKLVLDLSENFSGELPESTLNSASLLTLDSSNPFGLPMLQCNPK 415  
DB 160 YPTDIL-----LPMKLGYNLKTG-LNRFLTWSRS-----DDPSGNGFSYTL-----ET 203  
QY 416 NTLQELVYKNNGFTGKIPPTLSNCSLVSLSLSPNYLSGTPISLSLSKRLDKLMLNM 475  
DB 204 QSLPEFYLSRNF-----PMHRS-----GFWNGI 227  
QY 476 LEGEIPQELMVKTLFLILLDPNDLGEIPLSGLSNCTNLNMTLSNNRUTGELPKVIGRL 535  
DB 228 RPSGIPED-----QLSTYVYVNFLENNEEV-AYTFRTNNSFSY-----RLTISSEGFQGL 278  
QY 536 ENLAILKLISNNSFSGNIPDELGD-----CRSLIMDLWT-----NLFNCTIPAAAFKQSGKIA 588  
DB 279 TWYSIRIMRPFMSPPVPCDITYMCGPFAVCDVNTSPVCNCIOGFNFRNIQOWDORVW 338  
QY 589 ANFLAGKRYVYIKNDG--MKKECHAGNLEFQGISPOLNRLSTR--NPGNITSRYVG 643  
DB 339 AGGCIKRTQSCSGDGFTRMKRKKLPEITMAVVD--RSIGVACKKRCISDNCCTA----- 392  
QY 644 GHTSPPTPDN-----NGSMFLDMSYMLSGVIRKESNPFYFLNLGHNDISGSI--PDE 697  
DB 393 -----FANADIRNGG-----SGCVIWTFR 411  
QY 698 VGLRGLNILLDSNKLDRIPQMSALMTLLEIDLS--NNLSGPIPMGQEPFPPAKF 756  
DB 412 LEDIR-----NVATDAIDQ-----DLVRLAADIYAKRNASGKI----- 447  
QY 757 LNNPGLGYPVLRPCDSNADGYAHHORSHRRRPSLAGSVAMGLFSPVCIPLGLVGRE 816  
DB 448 -----ISLTGVSVLTLILWFCLM----- 466  
QY 817 MRKRKKKEAELEMYAEHGNSGDRPANTNMWKLGV-----KEALSTLAAFEKPKLKT 872  
DB 467 KRQKRAKXSAISI-----ANTQRNQLPMNEMVLSKRSRSEGEYKEELPLIE 517  
QY 873 FADLLATNGFINDLSISGGFGDYKAILKDGSAVAIKLIIHVGQDREPMEMETIG 932  
DB 518 METVAVATNPFSSCNKLTQGGFGIYKGRLLDKKEIAVKRLSKTSVGTDEPMNEVLIA 577  
QY 933 KIKHRLVLLGYCVGSERLVNEVMKYSLEVDLPDPKGGYVYKLSSTRKIAIGSAR 992  
DB 578 RLQHNILVGLCCLESGDKMLIYELENISLDSYLFQ-KYRSKLMWNERFDTNGVAR 636  
QY 993 GLAFIHNCSPIIHRDMKSSNVLLDENTLBARVDFPGMARLMSAMDTLSVSTLAGPGY 1052  
DB 637 GLLYLHODSRFPIIHRDLKVSNIILLDKMIPKISDFGMALIFERDETEANTMKVGVGGY 696  
QY 1053 VPPEYVQFRGCTKQDVYSYGVVLELLITGK--RPTDSPFGDNNILVGVV-----KQHAKT 1106  
DB 697 NSPEYAMVGIKSEKSDVFSFGVIVLEIVSGKNGKGFNYLDY-ENDDLSYVSWKXGGRAL 755  
QY 1107 RISDVDEELMKEDPAL--EIELQHLKVAVACLDPRAMRPTVMYMMF 1155  
DB 756 EIVDPAVIVDSLSQPSIFQPOEVAKICIGILCYQVLAELHRRPMSSVWVWF 806

RESULT 11  
ID D100 ARATH STANDARD; PRT; 372 AA.  
AC Q00874; O9LHKO;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DNA-damage-repair/coleraction protein DRT100 precursor.  
GN DRT100 OR AT3G12610 OR MMF12.5 OR T2E22.8 OR T2E22\_107.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

Query Match	Best Local Similarity	Matches	Score	DB 1	Length	DB 2	Score	DB 3	Length	DB 4	Score	DB 5	Length	DB 6	Score	DB 7	Length	DB 8	Score	DB 9	Length	DB 10	Score	DB 11	Length	DB 12	Score	DB 13	Length	DB 14	Score	DB 15	Length	DB 16	Score	DB 17	Length	DB 18	Score	DB 19	Length	DB 20	Score	DB 21	Length	DB 22	Score	DB 23	Length	DB 24	Score	DB 25	Length	DB 26	Score	DB 27	Length	DB 28	Score	DB 29	Length	DB 30	Score	DB 31	Length	DB 32	Score	DB 33	Length	DB 34	Score	DB 35	Length	DB 36	Score	DB 37	Length	DB 38	Score	DB 39	Length	DB 40	Score	DB 41	Length	DB 42	Score	DB 43	Length	DB 44	Score	DB 45	Length	DB 46	Score	DB 47	Length	DB 48	Score	DB 49	Length	DB 50	Score	DB 51	Length	DB 52	Score	DB 53	Length	DB 54	Score	DB 55	Length	DB 56	Score	DB 57	Length	DB 58	Score	DB 59	Length	DB 60	Score	DB 61	Length	DB 62	Score	DB 63	Length	DB 64	Score	DB 65	Length	DB 66	Score	DB 67	Length	DB 68	Score	DB 69	Length	DB 70	Score	DB 71	Length	DB 72	Score	DB 73	Length	DB 74	Score	DB 75	Length	DB 76	Score	DB 77	Length	DB 78	Score	DB 79	Length	DB 80	Score	DB 81	Length	DB 82	Score	DB 83	Length	DB 84	Score	DB 85	Length	DB 86	Score	DB 87	Length	DB 88	Score	DB 89	Length	DB 90	Score	DB 91	Length	DB 92	Score	DB 93	Length	DB 94	Score	DB 95	Length	DB 96	Score	DB 97	Length	DB 98	Score	DB 99	Length	DB 100	Score	DB 101	Length	DB 102	Score	DB 103	Length	DB 104	Score	DB 105	Length	DB 106	Score	DB 107	Length	DB 108	Score	DB 109	Length	DB 110	Score	DB 111	Length	DB 112	Score	DB 113	Length	DB 114	Score	DB 115	Length	DB 116	Score	DB 117	Length	DB 118	Score	DB 119	Length	DB 120	Score	DB 121	Length	DB 122	Score	DB 123	Length	DB 124	Score	DB 125	Length	DB 126	Score	DB 127	Length	DB 128	Score	DB 129	Length	DB 130	Score	DB 131	Length	DB 132	Score	DB 133	Length	DB 134	Score	DB 135	Length	DB 136	Score	DB 137	Length	DB 138	Score	DB 139	Length	DB 140	Score	DB 141	Length	DB 142	Score	DB 143	Length	DB 144	Score	DB 145	Length	DB 146	Score	DB 147	Length	DB 148	Score	DB 149	Length	DB 150	Score	DB 151	Length	DB 152	Score	DB 153	Length	DB 154	Score	DB 155	Length	DB 156	Score	DB 157	Length	DB 158	Score	DB 159	Length	DB 160	Score	DB 161	Length	DB 162	Score	DB 163	Length	DB 164	Score	DB 165	Length	DB 166	Score	DB 167	Length	DB 168	Score	DB 169	Length	DB 170	Score	DB 171	Length	DB 172	Score	DB 173	Length	DB 174	Score	DB 175	Length	DB 176	Score	DB 177	Length	DB 178	Score	DB 179	Length	DB 180	Score	DB 181	Length	DB 182	Score	DB 183	Length	DB 184	Score	DB 185	Length	DB 186	Score	DB 187	Length	DB 188	Score	DB 189	Length	DB 190	Score	DB 191	Length	DB 192	Score	DB 193	Length	DB 194	Score	DB 195	Length	DB 196	Score	DB 197	Length	DB 198	Score	DB 199	Length	DB 200	Score	DB 201	Length	DB 202	Score	DB 203	Length	DB 204	Score	DB 205	Length	DB 206	Score	DB 207	Length	DB 208	Score	DB 209	Length	DB 210	Score	DB 211	Length	DB 212	Score	DB 213	Length	DB 214	Score	DB 215	Length	DB 216	Score	DB 217	Length	DB 218	Score	DB 219	Length	DB 220	Score	DB 221	Length	DB 222	Score	DB 223	Length	DB 224	Score	DB 225	Length	DB 226	Score	DB 227	Length	DB 228	Score	DB 229	Length	DB 230	Score	DB 231	Length	DB 232	Score	DB 233	Length	DB 234	Score	DB 235	Length	DB 236	Score	DB 237	Length	DB 238	Score	DB 239	Length	DB 240	Score	DB 241	Length	DB 242	Score	DB 243	Length	DB 244	Score	DB 245	Length	DB 246	Score	DB 247	Length	DB 248	Score	DB 249	Length	DB 250	Score	DB 251	Length	DB 252	Score	DB 253	Length	DB 254	Score	DB 255	Length	DB 256	Score	DB 257	Length	DB 258	Score	DB 259	Length	DB 260	Score	DB
-------------	-----------------------	---------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	----



RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98284030; PubMed=9618511;  
 RA Selters L.M., Schutzman J.L., Borland C.Z., Stern M.J.;  
 RT "Soc-2 encodes a leucine-rich repeat protein implicated in fibroblast  
 growth factor receptor signaling.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:6903-6908(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98337190; PubMed=9674433;  
 RA Steurich D.S., Sun Q., Han M.;  
 RT "Soc-8, a conserved Ras-binding protein with leucine-rich repeats,  
 positively regulates Ras-mediated signaling in C. elegans.";  
 RL Cell 94:119-130(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99156230; PubMed=10048485;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohta O.;  
 RT "Prediction of the coding sequences of unidentified human genes.  
 III. The complete sequences of 100 new cDNA clones from brain which  
 code for large proteins in vitro.";  
 RL DNA Res. 5:355-364(1998).  
 CC -1 SUBUNIT: SPECIFICALLY BINDS K-RAS AND N-RAS BUT NOT H-RAS IN  
 VITRO.  
 CC -1 SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; AF054828; AAC35698.1; -;  
 DR EMBL; AF068820; AAC39856.1; -;  
 DR EMBL; AB020669; BAA74885.1; -;  
 DR Genew; HGNC:15454; SHOC2.  
 DR MIM; 602775; -;  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003591; LRR\_typ.  
 DR Pfam; PF00560; LRR\_17.  
 DR PRINTS; PRO0019; LEURICRPT.  
 DR SMART; SM00369; LRR\_TYP; 12.  
 KW Repeat; Leucine-rich repeat.  
 FT REPEAT 99 122 LRR 1.  
 FT REPEAT 123 145 LRR 2.  
 FT REPEAT 146 168 LRR 3.  
 FT REPEAT 169 191 LRR 4.  
 FT REPEAT 193 214 LRR 5.  
 FT REPEAT 215 237 LRR 6.  
 FT REPEAT 239 260 LRR 7.  
 FT REPEAT 262 283 LRR 8.  
 FT REPEAT 284 306 LRR 9.  
 FT REPEAT 307 329 LRR 10.  
 FT REPEAT 331 353 LRR 11.  
 FT REPEAT 354 377 LRR 12.  
 FT REPEAT 379 400 LRR 13.  
 FT REPEAT 401 424 LRR 14.  
 FT REPEAT 425 447 LRR 15.  
 FT REPEAT 448 471 LRR 16.  
 FT REPEAT 473 493 LRR 17.  
 FT REPEAT 494 516 LRR 18.  
 FT REPEAT 518 540 LRR 19.  
 FT REPEAT 541 563 LRR 20.  
 FT REPEAT 565 582 LRR 21.  
 FT REPEAT 582 604 LRR 22.  
 FT REPEAT 605 627 LRR 23.  
 FT REPEAT 628 650 LRR 24.  
 FT REPEAT 651 673 LRR 25.  
 FT REPEAT 674 696 LRR 26.  
 FT REPEAT 697 719 LRR 27.  
 FT REPEAT 720 742 LRR 28.  
 FT REPEAT 743 765 LRR 29.  
 FT REPEAT 766 788 LRR 30.  
 FT REPEAT 789 811 LRR 31.  
 FT REPEAT 812 834 LRR 32.  
 FT REPEAT 835 857 LRR 33.  
 FT REPEAT 858 880 LRR 34.  
 FT REPEAT 881 903 LRR 35.  
 FT REPEAT 904 926 LRR 36.  
 FT REPEAT 927 949 LRR 37.  
 FT REPEAT 950 972 LRR 38.  
 FT REPEAT 973 995 LRR 39.  
 FT REPEAT 996 1018 LRR 40.  
 FT REPEAT 1019 1041 LRR 41.  
 FT REPEAT 1042 1064 LRR 42.  
 FT REPEAT 1065 1087 LRR 43.  
 FT REPEAT 1088 1110 LRR 44.  
 FT REPEAT 1111 1133 LRR 45.  
 FT REPEAT 1134 1156 LRR 46.  
 FT REPEAT 1157 1179 LRR 47.  
 FT REPEAT 1180 1202 LRR 48.  
 FT REPEAT 1203 1225 LRR 49.  
 FT REPEAT 1226 1248 LRR 50.  
 FT REPEAT 1249 1271 LRR 51.  
 FT REPEAT 1272 1294 LRR 52.  
 FT REPEAT 1295 1317 LRR 53.  
 FT REPEAT 1318 1340 LRR 54.  
 FT REPEAT 1341 1363 LRR 55.  
 FT REPEAT 1364 1386 LRR 56.  
 FT REPEAT 1387 1409 LRR 57.  
 FT REPEAT 1410 1432 LRR 58.  
 FT REPEAT 1433 1455 LRR 59.  
 FT REPEAT 1456 1478 LRR 60.  
 FT REPEAT 1479 1501 LRR 61.  
 FT REPEAT 1502 1524 LRR 62.  
 FT REPEAT 1525 1547 LRR 63.  
 FT REPEAT 1548 1570 LRR 64.  
 FT REPEAT 1571 1593 LRR 65.  
 FT REPEAT 1594 1616 LRR 66.  
 FT REPEAT 1617 1639 LRR 67.  
 FT REPEAT 1640 1662 LRR 68.  
 FT REPEAT 1663 1685 LRR 69.  
 FT REPEAT 1686 1708 LRR 70.  
 FT REPEAT 1709 1731 LRR 71.  
 FT REPEAT 1732 1754 LRR 72.  
 FT REPEAT 1755 1777 LRR 73.  
 FT REPEAT 1778 1800 LRR 74.  
 FT REPEAT 1801 1823 LRR 75.  
 FT REPEAT 1824 1846 LRR 76.  
 FT REPEAT 1847 1869 LRR 77.  
 FT REPEAT 1870 1892 LRR 78.  
 FT REPEAT 1893 1915 LRR 79.  
 FT REPEAT 1916 1938 LRR 80.  
 FT REPEAT 1939 1961 LRR 81.  
 FT REPEAT 1962 1984 LRR 82.  
 FT REPEAT 1985 2007 LRR 83.  
 FT REPEAT 2008 2030 LRR 84.  
 FT REPEAT 2031 2053 LRR 85.  
 FT REPEAT 2054 2076 LRR 86.  
 FT REPEAT 2077 2099 LRR 87.  
 FT REPEAT 2100 2122 LRR 88.  
 FT REPEAT 2123 2145 LRR 89.  
 FT REPEAT 2146 2168 LRR 90.  
 FT REPEAT 2169 2191 LRR 91.  
 FT REPEAT 2192 2214 LRR 92.  
 FT REPEAT 2215 2237 LRR 93.  
 FT REPEAT 2238 2260 LRR 94.  
 FT REPEAT 2261 2283 LRR 95.  
 FT REPEAT 2284 2306 LRR 96.  
 FT REPEAT 2307 2329 LRR 97.  
 FT REPEAT 2330 2352 LRR 98.  
 FT REPEAT 2353 2375 LRR 99.  
 FT REPEAT 2376 2398 LRR 100.  
 FT REPEAT 2399 2421 LRR 101.  
 FT REPEAT 2422 2444 LRR 102.  
 FT REPEAT 2445 2467 LRR 103.  
 FT REPEAT 2468 2490 LRR 104.  
 FT REPEAT 2491 2513 LRR 105.  
 FT REPEAT 2514 2536 LRR 106.  
 FT REPEAT 2537 2559 LRR 107.  
 FT REPEAT 2560 2582 LRR 108.  
 FT REPEAT 2583 2605 LRR 109.  
 FT REPEAT 2606 2628 LRR 110.  
 FT REPEAT 2629 2651 LRR 111.  
 FT REPEAT 2652 2674 LRR 112.  
 FT REPEAT 2675 2697 LRR 113.  
 FT REPEAT 2698 2720 LRR 114.  
 FT REPEAT 2721 2743 LRR 115.  
 FT REPEAT 2744 2766 LRR 116.  
 FT REPEAT 2767 2789 LRR 117.  
 FT REPEAT 2790 2812 LRR 118.  
 FT REPEAT 2813 2835 LRR 119.  
 FT REPEAT 2836 2858 LRR 120.  
 FT REPEAT 2859 2881 LRR 121.  
 FT REPEAT 2882 2904 LRR 122.  
 FT REPEAT 2905 2927 LRR 123.  
 FT REPEAT 2928 2950 LRR 124.  
 FT REPEAT 2951 2973 LRR 125.  
 FT REPEAT 2974 2996 LRR 126.  
 FT REPEAT 2997 3019 LRR 127.  
 FT REPEAT 3020 3042 LRR 128.  
 FT REPEAT 3043 3065 LRR 129.  
 FT REPEAT 3066 3088 LRR 130.  
 FT REPEAT 3089 3111 LRR 131.  
 FT REPEAT 3112 3134 LRR 132.  
 FT REPEAT 3135 3157 LRR 133.  
 FT REPEAT 3158 3180 LRR 134.  
 FT REPEAT 3181 3203 LRR 135.  
 FT REPEAT 3204 3226 LRR 136.  
 FT REPEAT 3227 3249 LRR 137.  
 FT REPEAT 3250 3272 LRR 138.  
 FT REPEAT 3273 3295 LRR 139.  
 FT REPEAT 3296 3318 LRR 140.  
 FT REPEAT 3319 3341 LRR 141.  
 FT REPEAT 3342 3364 LRR 142.  
 FT REPEAT 3365 3387 LRR 143.  
 FT REPEAT 3388 3410 LRR 144.  
 FT REPEAT 3411 3433 LRR 145.  
 FT REPEAT 3434 3456 LRR 146.  
 FT REPEAT 3457 3479 LRR 147.  
 FT REPEAT 3480 3502 LRR 148.  
 FT REPEAT 3503 3525 LRR 149.  
 FT REPEAT 3526 3548 LRR 150.  
 FT REPEAT 3549 3571 LRR 151.  
 FT REPEAT 3572 3594 LRR 152.  
 FT REPEAT 3595 3617 LRR 153.  
 FT REPEAT 3618 3640 LRR 154.  
 FT REPEAT 3641 3663 LRR 155.  
 FT REPEAT 3664 3686 LRR 156.  
 FT REPEAT 3687 3709 LRR 157.  
 FT REPEAT 3710 3732 LRR 158.  
 FT REPEAT 3733 3755 LRR 159.  
 FT REPEAT 3756 3778 LRR 160.  
 FT REPEAT 3779 3801 LRR 161.  
 FT REPEAT 3802 3824 LRR 162.  
 FT REPEAT 3825 3847 LRR 163.  
 FT REPEAT 3848 3870 LRR 164.  
 FT REPEAT 3871 3893 LRR 165.  
 FT REPEAT 3894 3916 LRR 166.  
 FT REPEAT 3917 3939 LRR 167.  
 FT REPEAT 3940 3962 LRR 168.  
 FT REPEAT 3963 3985 LRR 169.  
 FT REPEAT 3986 4008 LRR 170.  
 FT REPEAT 4009 4031 LRR 171.  
 FT REPEAT 4032 4054 LRR 172.  
 FT REPEAT 4055 4076 LRR 173.  
 FT REPEAT 4077 4099 LRR 174.  
 FT REPEAT 4100 4122 LRR 175.  
 FT REPEAT 4123 4144 LRR 176.  
 FT REPEAT 4145 4167 LRR 177.  
 FT REPEAT 4168 4190 LRR 178.  
 FT REPEAT 4191 4212 LRR 179.  
 FT REPEAT 4213 4235 LRR 180.  
 FT REPEAT 4236 4257 LRR 181.  
 FT REPEAT 4258 4280 LRR 182.  
 FT REPEAT 4281 4303 LRR 183.  
 FT REPEAT 4304 4325 LRR 184.  
 FT REPEAT 4326 4348 LRR 185.  
 FT REPEAT 4349 4370 LRR 186.  
 FT REPEAT 4371 4393 LRR 187.  
 FT REPEAT 4394 4415 LRR 188.  
 FT REPEAT 4416 4438 LRR 189.  
 FT REPEAT 4439 4460 LRR 190.  
 FT REPEAT 4461 4482 LRR 191.  
 FT REPEAT 4483 4505 LRR 192.  
 FT REPEAT 4506 4527 LRR 193.  
 FT REPEAT 4528 4550 LRR 194.  
 FT REPEAT 4551 4572 LRR 195.  
 FT REPEAT 4573 4594 LRR 196.  
 FT REPEAT 4595 4617 LRR 197.  
 FT REPEAT 4618 4639 LRR 198.  
 FT REPEAT 4640 4661 LRR 199.  
 FT REPEAT 4662 4684 LRR 200.  
 FT REPEAT 4685 4706 LRR 201.  
 FT REPEAT 4707 4728 LRR 202.  
 FT REPEAT 4729 4751 LRR 203.  
 FT REPEAT 4752 4773 LRR 204.  
 FT REPEAT 4774 4796 LRR 205.  
 FT REPEAT 4797 4818 LRR 206.  
 FT REPEAT 4819 4840 LRR 207.  
 FT REPEAT 4841 4863 LRR 208.  
 FT REPEAT 4864 4885 LRR 209.  
 FT REPEAT 4886 4908 LRR 210.  
 FT REPEAT 4909 4930 LRR 211.  
 FT REPEAT 4931 4953 LRR 212.  
 FT REPEAT 4954 4975 LRR 213.  
 FT REPEAT 4976 4998 LRR 214.  
 FT REPEAT 4999 5020 LRR 215.  
 FT REPEAT 5021 5043 LRR 216.  
 FT REPEAT 5044 5065 LRR 217.  
 FT REPEAT 5066 5088 LRR 218.  
 FT REPEAT 5089 5110 LRR 219.  
 FT REPEAT 5111 5132 LRR 220.  
 FT REPEAT 5133 5155 LRR 221.  
 FT REPEAT 5156 5177 LRR 222.  
 FT REPEAT 5178 5199 LRR 223.  
 FT REPEAT 5200 5221 LRR 224.  
 FT REPEAT 5222 5244 LRR 225.  
 FT REPEAT 5245 5266 LRR 226.  
 FT REPEAT 5267 5289 LRR 227.  
 FT REPEAT 5290 5311 LRR 228.  
 FT REPEAT 5312 5334 LRR 229.  
 FT REPEAT 5335 5356 LRR 230.  
 FT REPEAT 5357 5378 LRR 231.  
 FT REPEAT 5379 5400 LRR 232.  
 FT REPEAT 5401 5423 LRR 233.  
 FT REPEAT 5424 5445 LRR 234.  
 FT REPEAT 5446 5468 LRR 235.  
 FT REPEAT 5469 5490 LRR 236.  
 FT REPEAT 5491 5512 LRR 237.  
 FT REPEAT 5513 5534 LRR 238.  
 FT REPEAT 5535 5557 LRR 239.  
 FT REPEAT 5558 5579 LRR 240.  
 FT REPEAT 5580 5602 LRR 241.  
 FT REPEAT 5603 5624 LRR 242.  
 FT REPEAT 5625 5647 LRR 243.  
 FT REPEAT 5648 5669 LRR 244.  
 FT REPEAT 5670 5692 LRR 245.  
 FT REPEAT 5693 5714 LRR 246.  
 FT REPEAT 5715 5737 LRR 247.  
 FT REPEAT 5738 5759 LRR 248.  
 FT REPEAT 5760 5782 LRR 249.  
 FT REPEAT 5783 5804 LRR 250.  
 FT REPEAT 5805 5826 LRR 251.  
 FT REPEAT 5827 5849 LRR 252.  
 FT REPEAT 5850 5871 LRR 253.  
 FT REPEAT 5872 5894 LRR 254.  
 FT REPEAT 5895 5916 LRR 255.  
 FT REPEAT 5917 5938 LRR 256.  
 FT REPEAT 5939 5960 LRR 257.  
 FT REPEAT 5961 5983 LRR 258.  
 FT REPEAT 5984 6005 LRR 259.  
 FT REPEAT 6006 6027 LRR 260.  
 FT REPEAT 6028 6050 LRR 261.  
 FT REPEAT 6051 6071 LRR 262.  
 FT REPEAT 6072 6094 LRR 263.  
 FT REPEAT 6095 6115 LRR 264.  
 FT REPEAT 6116 6138 LRR 265.  
 FT REPEAT 6139 6160 LRR 266.  
 FT REPEAT 6161 6182 LRR 267.  
 FT REPEAT 6183 6204 LRR 268.  
 FT REPEAT 6205 6227 LRR 269.  
 FT REPEAT 6228 6249 LRR 270.  
 FT REPEAT 6250 6271 LRR 271.  
 FT REPEAT 6272 6294 LRR 272.  
 FT REPEAT 6295 6315 LRR 273.  
 FT REPEAT 6316 6338 LRR 274.  
 FT REPEAT 6339 6360 LRR 275.  
 FT REPEAT 6361 6382 LRR 276.  
 FT REPEAT 6383 6405 LRR 277.  
 FT REPEAT 6406 6426 LRR 278.  
 FT REPEAT 6427 6448 LRR 279.  
 FT REPEAT 6449 6470 LRR 280.  
 FT REPEAT 6471 6492 LRR 281.  
 FT REPEAT 6493 6514 LRR 282.  
 FT REPEAT 6515 6536 LRR 283.  
 FT REPEAT 6537 6558 LRR 284.  
 FT REPEAT 6559 6580 LRR 285.  
 FT REPEAT 6581 6602 LRR 286.  
 FT REPEAT 6603 6624 LRR 287.  
 FT REPEAT 6625 6645 LRR 288.  
 FT REPEAT 6646 6667 LRR 289.  
 FT REPEAT 6668 6689 LRR 290.  
 FT REPEAT 6690 6711 LRR 291.  
 FT REPEAT 6712 6732 LRR 292.  
 FT REPEAT 6733 6754 LRR 293.  
 FT REPEAT 6755 6775 LRR 294.  
 FT REPEAT 6776 6797 LRR 295.  
 FT REPEAT 6798 6818 LRR 296.  
 FT REPEAT 6819 6840 LRR 297.  
 FT REPEAT 6841 6861 LRR 298.  
 FT REPEAT 6862 6883 LRR 299.  
 FT REPEAT 6884 6904 LRR 300.  
 FT REPEAT 6905 6926 LRR 301.  
 FT REPEAT 6927 6947 LRR 302.  
 FT REPEAT 6948 6968 LRR 303.  
 FT REPEAT 6969 6990 LRR 304.  
 FT REPEAT 6991 7010 LRR 305.  
 FT REPEAT 7011 7032 LRR 306.  
 FT REPEAT 7033 7053 LRR 307.  
 FT REPEAT 7054 7074 LRR 308.  
 FT REPEAT 7075 7095 LRR 309.  
 FT REPEAT 7096 7116 LRR 310.  
 FT REPEAT 7117 7137 LRR 311.  
 FT REPEAT 7138 7158 LRR 312.  
 FT REPEAT 7159 7179 LRR 313.  
 FT REPEAT 7180 7200 LRR 314.  
 FT REPEAT 7201 7221 LRR 315.  
 FT REPEAT 7222 7242 LRR 316.  
 FT REPEAT 7243 7263 LRR 317.  
 FT REPEAT 7264 7284 LRR 318.  
 FT REPEAT 7285 7305 LRR 319.  
 FT REPEAT 7306 7326 LRR 320.  
 FT REPEAT 7327 7347 LRR 321.  
 FT REPEAT 7348 7368 LRR 322.  
 FT REPEAT 7369 7389 LRR 323.  
 FT REPEAT 7390 7410 LRR 324.  
 FT REPEAT 7411 7430 LRR 325.  
 FT REPEAT 7431 7451 LRR 326.  
 FT REPEAT 7452 7472 LRR 327.  
 FT REPEAT 7473 7492 LRR 328.  
 FT REPEAT 7493 7513 LRR 329.  
 FT REPEAT 7514 7533 LRR 330.  
 FT REPEAT 7534 7554 LRR 331.  
 FT REPEAT 7555 7574 LRR 332.  
 FT REPEAT 7575 7594 LRR 333.  
 FT REPEAT 7595 7614 LRR 334.  
 FT REPEAT 7615 7634 LRR 335.  
 FT REPEAT 7635 7654 LRR 336.  
 FT REPEAT 7655 7674 LRR 337.  
 FT REPEAT 7675 7694 LRR 338.  
 FT REPEAT 7695 7713 LRR 339.  
 FT REPEAT 7714 7733 LRR 340.  
 FT REPEAT 7734 7753 LRR 341.  
 FT REPEAT 7754 7772 LRR 342.  
 FT REPEAT 7773 7792 LRR 343.  
 FT REPEAT 7793 7811 LRR 344.  
 FT REPEAT 7812 7830 LRR 345.  
 FT REPEAT 7831 7849 LRR 346.  
 FT REPEAT 7850 7868 LRR 347.  
 FT REPEAT 7869 7887 LRR 348.  
 FT REPEAT 7888 7906 LRR 349.  
 FT REPEAT 7907 7924 LRR 350.  
 FT REPEAT 7925 7943 LRR 351.  
 FT REPEAT 7944 7961 LRR 352.  
 FT REPEAT 7962 7980 LRR 353.  
 FT REPEAT 7981 7998 LRR 354.  
 FT REPEAT 7999 8016 LRR 355.  
 FT REPEAT 8017 8034 LRR 356.  
 FT REPEAT 8035 8052 LRR 357.  
 FT REPEAT 8053 8070 LRR 358.  
 FT REPEAT 8071 8088 LRR 359.  
 FT REPEAT 8089 8106 LRR 360.  
 FT REPEAT 8107 8124 LRR 361.  
 FT REPEAT 8125 8142 LRR 362.  
 FT REPEAT 8143 8160 LRR 363.  
 FT REPEAT 8161 8178 LRR 364.  
 FT REPEAT 8179 8196 LRR 365.  
 FT REPEAT 8197 8214 LRR 366.  
 FT REPEAT 8215 8232 LRR 367.  
 FT REPEAT 8233 8250 LRR 368.  
 FT REPEAT 8251 8268 LRR 369.  
 FT REPEAT 8269 8286 LRR 370.  
 FT REPEAT 8287 8304 LRR 371.  
 FT REPEAT 8305 8322 LRR 372.  
 FT REPEAT 8323 8340 LRR 373.  
 FT REPEAT 8341 8358 LRR 374.  
 FT REPEAT 8359 8376 LRR 375.  
 FT REPEAT 8377 8394 LRR 376.  
 FT REPEAT 8395 8412 LRR 377.  
 FT REPEAT 8413 8430 LRR 378.  
 FT REPEAT 8431 8448 LRR 379.  
 FT REPEAT 8449 8466 LRR 380.  
 FT REPEAT 8467 8484 LRR 381.  
 FT REPEAT 8485 8502 LRR 382.  
 FT REPEAT 8503 8520 LRR 383.  
 FT REPEAT 8521 8538 LRR 384.  
 FT REPEAT 8539 8556 LRR 385.  
 FT REPEAT 8557 8574 LRR 386.  
 FT REPEAT 8575 8592 LRR 387.  
 FT REPEAT 8593 8610 LRR 388.  
 FT REPEAT 8611 8628 LRR 389.  
 FT REPEAT 8629 8646 LRR 390.  
 FT REPEAT 8647 8664 LRR 391.  
 FT REPEAT 8665 8682 LRR 392.  
 FT REPEAT 8683 8700 LRR 393.  
 FT REPEAT 8701 8718 LRR 394.  
 FT REPEAT 8719 8736 LRR 395.  
 FT REPEAT 8737 8754 LRR 396.  
 FT REPEAT 8755 8772 LRR 397.  
 FT REPEAT 8773 8790 LRR 398.  
 FT REPEAT 8791 8808 LRR 399.  
 FT REPEAT 8809 8826 LRR 400.  
 FT REPEAT 8827 8844 LRR 401.  
 FT REPEAT 8845 8862 LRR 402.  
 FT REPEAT 8863 8880 LRR 403.  
 FT REPEAT 8881 8898 LRR 404.  
 FT REPEAT 8899 8916 LRR 405.  
 FT REPEAT 8917 8934 LRR 406.  
 FT REPEAT 8935 8952 LRR 407.  
 FT REPEAT 8953 8970 LRR 408.  
 FT REPEAT 8971 8988 LRR 409.  
 FT REPEAT 8989 9006 LRR 410.  
 FT REPEAT 9007 9024 LRR 411.  
 FT REPEAT 9025 9042 LRR 412.  
 FT REPEAT 9043 9060 LRR 413.  
 FT REPEAT 9061 9078 LRR 414.  
 FT REPEAT 9079 9096 LRR 415.  
 FT REPEAT 9097 9114 LRR 416.  
 FT REPEAT 9115 9132 LRR 417.  
 FT REPEAT 9133 9150 LRR 418.  
 FT REPEAT 9151 9168 LRR 419.  
 FT REPEAT 9169 9186 LRR 420.  
 FT REPEAT 9187 9204 LRR 421.  
 FT REPEAT 9205 9222 LRR 422.  
 FT REPEAT 9223 9240 LRR 423.  
 FT REPEAT 9241 9258 LRR 424.  
 FT REPEAT 9259 9276 LRR 425.  
 FT REPEAT 9277 9294 LRR 426.  
 FT REPEAT 9295 9312 LRR 427.  
 FT REPEAT 9313 9330 LRR 428.  
 FT REPEAT 9331 9348 LRR 429.  
 FT REPEAT 9349 9366 LRR 430.  
 FT REPEAT 9367 9384 LRR 431.  
 FT REPEAT 9385 9402 LRR 432.  
 FT REPEAT 9403 9420 LRR 433.  
 FT REPEAT 9421 9438 LRR 434.  
 FT REPEAT 9439 9456 LRR 435.  
 FT REPEAT 9457 9474 LRR 436.  
 FT REPEAT 9475 9492 LRR 437.  
 FT REPEAT 9493 9510 LRR 438.  
 FT REPEAT 9511 9528 LRR 439.  
 FT REPEAT 9529 9546 LRR 440.  
 FT REPEAT 9547 9564 LRR 441.  
 FT REPEAT 9565 9582 LRR 442.  
 FT REPEAT 9583 9600 LRR 443.  
 FT REPEAT 9601 9618 LRR 444.  
 FT REPEAT 9619 9636 LRR 445.  
 FT REPEAT 9637 9654 LRR 446.  
 FT REPEAT 9655 9672 LRR 447.  
 FT REPEAT 9673 9690 LRR 448.  
 FT REPEAT 9691 9708 LRR 449.  
 FT REPEAT 9709 9726 LRR 450.  
 FT REPEAT 9727 9744 LRR 451.  
 FT REPEAT 9745 9762 LRR 452.  
 FT REPEAT 9763 9780 LRR 453.  
 FT REPEAT 9781 9798 LRR 454.  
 FT REPEAT 9799 9816 LRR 455.  
 FT REPEAT 9817 9834 LRR 456.  
 FT REPEAT 9835 9852 LRR 457.  
 FT REPEAT 9853 9870 LRR 458.  
 FT REPEAT 9871 9888 LRR 459.  
 FT REPEAT 9889 9906 LRR 460.  
 FT REPEAT 9907 9924 LRR 461.  
 FT REPEAT 9925 9942 LRR 462.  
 FT REPEAT 9943 9960 LRR 463.  
 FT REPEAT 9961 9978 LRR 464.  
 FT REPEAT 9979 9996 LRR 465.  
 FT REPEAT 9997 10000 LRR 466.  
 FT REPEAT 10001 10000 LRR 467.  
 FT REPEAT 10001 10000 LRR 468.  
 FT REPEAT 10001 10000 LRR 469.  
 FT REPEAT 10001 10000 LRR 470.  
 FT REPEAT 10001 10000 LRR 471.  
 FT REPEAT 10001 10000 LRR 472.  
 FT REPEAT 10001 10000 LRR 473.  
 FT REPEAT 10001 10000 LRR 474.  
 FT REPEAT 10001 10000 LRR 475.  
 FT REPEAT 10001 10000 LRR 476.  
 FT REPEAT 10001 10000 LRR 477.  
 FT REPEAT



## PELLE SUBFAMILY

-1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L08476; AAA28750.1; -  
 DR EMBL; AE003760; AAF56686.1; -  
 DR PIR; A45775; A45775  
 DR PDB; 1D2Z; 29-NOV-99.  
 DR Pfam; PF000010441; p11.  
 DR InterPro; IPR000488; Death.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00531; death; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SMO0005; DEATH; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.  
 DR TRANSFAC; Serine/threonine-protein kinase; ATP-binding;  
 KW 3D-structure.  
 FT DOMAIN 55 121  
 FT NP\_BIND 213 499 DEATH.  
 FT BINDING 219 227 PROTEIN KINASE.  
 FT ACT\_SITE 240 240 ATP (BY SIMILARITY).  
 FT MUTAGEN 346 346  
 FT MUTAGEN 240 240  
 FT MUTAGEN 346 346  
 FT MUTAGEN 350 350  
 FT SEQUENCE 501 AA; 56160 MW; 4829E2B40ACB81A8 CRC64;

Query Match 6.2%; Score 384.5; DB 1; Length 501;  
 Best Local Similarity 31.4%; Pred. No. 3.8e-15;  
 Matches 108; Conservative 75; Mismatches 132; Indels 29; Gaps 10;

QY 837 NSGRTNNTMKLTGKALSL-----INLAERK---PLRLTPADLIQATNGPHND 887  
 DB 158 SSGVSNNNRSTTATETEPSESLGNIHISTVORASLEIDVLELNATDMSPDN 217  
 QY 888 LIGSGGFDVYKAILKDGSAVAIKLIHSGGQDREFMA-----EMETGKIKRNLVP 941  
 DB 218 RLGGGGGDDVYRKWKQ-LDVALKVWYRSPNIDQKVELQGSYNELKTLNIRHDNIIA 276  
 QY 942 LIGYCKAGDELLVNEVMKXSLDEL-----QDPKGGVKKLSTRRKINAGSLAF 996  
 DB 277 LVGYSTIKGGPCVYVQIMKGSLEARLRAHKKQNPPLA---LTWQGRSISLGTAKGYTF 333  
 QY 997 LHNCSGPHIIHRDKMSNVLIDENLEARYSDGMAEL-MSANDTHLSVSTAGTPGVVP 1055  
 DB 334 LHTARGTPLIHGDIKAPNLLDQCLQPKIGDGLVREBGRKSLDAVENVKVFGRKIVLP 393  
 QY 1056 EYQSGRCSTGKDVSYGVVLELLTGKRPDPSDFGD--NNIVGVMQAHKALISVDP 1113  
 DB 394 EFRNFRQSLSTGVDSYFVILLEVFTGRQVTDVRENTKQLLDYKQQRNRMELE 453  
 QY 1114 PELMKEDPLETEL-LQHLKAVAVACLDRAVRRTPTVQVMAMFK 1156  
 DB 454 KKL-AAPMKELDMQCAIEAGHCTALDPODRPSMNAVLRKFE 496

RESULT 15  
 SHO2\_MOUSE STANDARD; PRT; 582 AA.  
 AC O88520;  
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Leucine-rich repeat protein SHOC-2 (Ras-binding protein Sur-8).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1] TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98337190; PubMed=9674433;  
 RA Sieburch D.S., Sun Q., Han W.;  
 RT "Sur-8, a conserved Ras-binding protein with leucine-rich repeats,  
 RT positively regulates Ras-mediated signaling in C. elegans.";  
 RL Cell 94:119-130(1998).  
 CC -1- SUBUNIT: SPECIFICALLY BINDS K-RAS AND N-RAS BUT NOT H-RAS IN  
 CC VIRO.  
 CC -1- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF068921; AAC40175.1; -  
 DR WGD; MGI:1927197; Shoc2.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR out.  
 DR InterPro; IPR003591; LRR typ.  
 DR Pfam; PF00560; LRR; 17.  
 DR PRINTS; PRO0019; LEUCRCHRT.  
 DR SMART; SMO0370; LRR; 8.  
 DR SMART; SMO0369; LRR typ; 5.  
 KW Repeat; Leucine-rich repeat.  
 FT REPEAT 99 122 LRR 1.  
 FT REPEAT 123 145 LRR 2.  
 FT REPEAT 146 168 LRR 3.  
 FT REPEAT 169 191 LRR 4.  
 FT REPEAT 193 214 LRR 5.  
 FT REPEAT 215 237 LRR 6.  
 FT REPEAT 239 260 LRR 7.  
 FT REPEAT 262 283 LRR 8.  
 FT REPEAT 284 306 LRR 9.  
 FT REPEAT 307 329 LRR 10.  
 FT REPEAT 331 353 LRR 11.  
 FT REPEAT 379 397 LRR 12.  
 FT REPEAT 401 424 LRR 13.  
 FT REPEAT 425 447 LRR 14.  
 FT REPEAT 448 471 LRR 15.  
 FT REPEAT 473 493 LRR 16.  
 FT REPEAT 494 516 LRR 17.  
 FT REPEAT 518 540 LRR 18.  
 FT REPEAT 541 563 LRR 19.  
 FT REPEAT 582 AA; 64947 MW; 62C5C082B3CF5F12 CRC64;

Query Match 6.1%; Score 378.5; DB 1; Length 582;  
 Best Local Similarity 28.9%; Pred. No. 1.1e-14;  
 Matches 151; Conservative 80; Mismatches 188; Indels 103; Gaps 24;

QY 69 CRDKVTSIDISKPLNMGSAVSSSLSTGLSEPLSNHSHNGSVSGFKGASLTSLD 128  
 DB 97 CRENSMRDLDSKRSIH-----LPPSVKELTQLTLYLXSNKIQSLPAVVGCLVNMIMTLA 152  
 QY 129 LSRNLSGPTTL-TISGSGSLKFLNVSNTL-DPPGVSGGLKLNSEVLDLSANST 186  
 DB 153 LSENSL---TSLPDSLDNKKLRMLDRHNKKLRIPSVV---YRLDSLTLTYLRFNIT 205  
 QY 187 GANVGVLSLD--GCGLKHLAISGNKISG-DVDVSRCVNLEFLVSSNNRSTGIPLD 243

Wed Mar 12 10:45:06 2003

us-09-823-394-2.rsp

Page 16

Db 206 -----TVEKDIKMLPYLSMISIRENKIKQIPAEIGELCNLITLVDVANHQ----- 249  
QY 244 CSALQHDIDISGNKLSGDFSAISCTETELKLNISNQFVGPIPPPLKSLQYLSLAENKE 303  
Db 250 ---LEHL-----PKEIGNCTQITNLDLQHNDLL----- 274  
QY 304 TGEIPDFLSGACDTLTGLDLSGNHFYGAVPPFPGSCSLLESLSLSSNPSGELP---MDT 360  
Db 275 --DLPDTI-GNLSSINRLGLRYNRL-SAIPLSLAKCSALELMLNENNTIS-TLPESLSS 329  
QY 361 LKRGKLVLDLSF-----NEFS-----GELPESLTNLSASLLITDLSN 400  
Db 330 LVKJNSLTLARNCFQLYVGGPSQSTIYSUNMEHNRINKIPGIFPRAKYLKLNKDN 389  
QY 401 NFSGPTLPNLCONPNTLOELYLQNNGFTGKIPTLSNCELYSLHSFYLSGTI PSSL 460  
Db 390 QUTS--LP-LDRGVTWSVLELNTATNQLT-KIPEVSGVLSLEVLILSNMLK-KLPHGL 444  
QY 461 GELSKLRDLKMLNMLEGEIPOELMYVKTLETLIDFNDLTGEIPSGLSNCTNIMWISLS 520  
445 GNLKRLRELDLEENKLE-SLPENIAYLKDQKLVITNNQLS-TLPBGIGHTNLTHTLGLG 502  
QY 521 NNRLTGEIPKWIQLENTLAILKLSNNSPESGNIPEDELGDGRSL 562  
Db 503 ENLLT-HLPEIGTLENLEIYLNPNLHSLPFELALCSKL 543

Search completed: March 10, 2003, 18:25:26  
Job time : 27.0143 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:11:50 ; Search time 34.2258 Seconds

(without alignment)  
3359.364 Million cell updates/sec

Title: US-09-823-394-2

Sequence: 1 MKTFSSFFLSVTTLFFSF...GPTTEMVDMISKEVBEGL 1196

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

PIR 73: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6140	99.3	1196	2	T09356
2	2623.5	42.4	1166	2	F96598
3	2509.5	40.6	1143	2	B84431
4	1608.5	26.0	1192	2	T48439
5	1419	23.0	1232	2	T05322
6	1380	22.3	1079	2	C96772
7	1370	22.2	1108	2	D84434
8	1367	22.1	1133	2	B86308
9	1364	22.1	1103	2	B84742
10	1357.5	22.0	1003	2	T05898
11	1349.5	21.8	1093	2	G96746
12	1340.5	21.7	1134	2	T04587
13	1335	21.6	1064	2	B86465
14	1313	21.2	1013	2	T10659
15	1311	21.2	999	1	S27756
16	1310	21.0	1002	2	T46033
17	1296.5	20.9	992	2	T05335
18	1293.5	20.9	992	2	T05335
19	1279.5	20.7	981	2	T50851
20	1273	20.6	1123	2	D96756
21	1257.5	20.3	996	2	F86410
22	1256.5	20.3	987	2	T50850
23	1230	19.9	1088	2	E86312
24	1237.5	19.9	976	2	B84659
25	1224.5	19.8	978	2	E96787
26	1222	19.8	932	2	T48489
27	1194.5	19.3	1027	2	B85089
28	1186.5	19.2	1120	2	B86479
29	1179	19.1	1029	2	T00712

30	1166.5	18.9	976	2	T05897	protein kinase hom
31	1147	18.6	1025	2	T45647	receptor protein k
32	1139	18.4	977	2	C96745	hypothetical prote
33	1124.5	18.2	991	2	T52400	receptor-like prot
34	1123	18.2	990	2	T03784	probable receptor
35	1109.5	17.9	1009	2	T45645	receptor kinase-li
36	1104	17.9	1029	2	T05050	protein kinase hom
37	1102.5	17.8	967	2	T48210	hypothetical prote
38	1094	17.7	964	2	T49308	hypothetical prote
39	1090	17.6	921	2	B86234	protein kinase xaz
40	1087	17.6	1025	1	A57676	probable receptor
41	1083	17.5	980	2	H84632	receptor-kinase 11
42	1081	17.5	1011	2	T45718	receptor-kinase 11
43	1069	17.3	988	2	T45717	hypothetical prote
44	1033	16.7	988	2	C96654	probable receptor
45	1031.5	16.7	890	2	E84846	

Query Match	99.3%	Score 6140;	DB 2;	Length 1196;
Best Local Similarity	99.4%;	Pred. No. 5.3e-272;		
Matches 1189;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;

QY	1	MKTFFSFFLSVTTLFFSFSSLSFQASPSQSLYREIHQLISFQVLPDKNLLPDMSSNNK	60	
DB	1	MKTFFSFFLSVTTLFFSFSSLSFQASPSQSLYREIHQLISFQVLPDKNLLPDMSSNNK	60	
QY	61	PCTFDGVTCDKDKVTSIDLSKSPINVFSAVSSLSLTGLTSLFTSNHSHNGSVGFC	120	
DB	61	PCTFDGVTCDKDKVTSIDLSKSPINVFSAVSSLSLTGLTSLFTSNHSHNGSVGFC	120	
QY	121	SASLTSLDLSNLSISGVTTTSLIGSCSGKLFNVSSNTLDPFGKVGGLKNSLEVDL	180	
DB	121	SASLTSLDLSNLSISGVTTTSLIGSCSGKLFNVSSNTLDPFGKVGGLKNSLEVDL	180	
QY	181	SANSISGANVGVWLSGCGELKHLAISGNKISGDVDSRCVNLLEFLDVSSNNFSGTGP	240	
DB	181	SANSISGANVGVWLSGCGELKHLAISGNKISGDVDSRCVNLLEFLDVSSNNFSGTGP	240	
QY	241	LGDCSALOHLDISGNKLSGDFSAISTCTELKLNLSNQFVGPPIPLPLKSLQYLSLAE	300	
DB	241	LGDCSALOHLDISGNKLSGDFSAISTCTELKLNLSNQFVGPPIPLPLKSLQYLSLAE	300	
QY	301	NKTFGEIPDLISGACDTLNLGLDSGNHFGVAVPPFGGCSLLSSTALSSNNFSGELPMDT	360	
DB	301	NKTFGEIPDLISGACDTLNLGLDSGNHFGVAVPPFGGCSLLSSTALSSNNFSGELPMDT	360	
QY	361	LAKRGKLVLDLSENFSGELPESTLNLASLTLTDLSSNNFSGTILNLTCONPKNTLOE	420	
DB	361	LAKRGKLVLDLSENFSGELPESTLNLASLTLTDLSSNNFSGTILNLTCONPKNTLOE	420	

## ALIGNMENTS

### RESULT 1

T09356  
Brassinosteroid-insensitive protein BR1 - Arabidopsis thaliana  
N/Alternate names: protein F23K16.30  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 24-Nov-1999  
C/Accession: T09356  
R/Byan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
Submitted to the Protein Sequence Database, June 1999  
A/Reference number: Z16652  
A/Accession: T09356  
A/Molecule type: DNA  
A/Residues: 1-1196 <BEV>  
A/Cross-references: EMBL:AL078620; GSPDB:GN00062; ATSP:F23K16.30  
A/Experimental source: cultivar Columbia; BAC clone F23K16  
C/Genetics:  
A/Map position: 4  
A/Map position: 4  
C/Superfamily: protein kinase Xaz1; leucine-rich alpha-2-glycoprotein repeat homology; p

QY 421 LYLONNGFTGKIPTLLSNCSSELYSLHSFNYSSTIPSSLSGLSKLRDLKLMNMBEGEI 480  
 Db 421 LYLONNGFTGKIPTLLSNCSSELYSLHSFNYSSTIPSSLSGLSKLRDLKLMNMBEGEI 480  
 QY 481 POELMYVKTLETLIIDFNDLGEIPSGLSNCTNLMWISLNNRKTGEIPKMTGRLENLAI 540  
 Db 481 POELMYVKTLETLIIDFNDLGEIPSGLSNCTNLMWISLNNRKTGEIPKMTGRLENLAI 540  
 QY 541 LKLSNNSSEGNTPDELGDGRSLIIMDLNTNLFNQTI PAAMFQSGKIAANFIAGRRYYVI 600  
 Db 541 LKLSNNSSEGNTPDELGDGRSLIIMDLNTNLFNQTI PAAMFQSGKIAANFIAGRRYYVI 600  
 QY 541 LKLSNNSSEGNTPDELGDGRSLIIMDLNTNLFNQTI PAAMFQSGKIAANFIAGRRYYVI 600  
 Db 541 LKLSNNSSEGNTPDELGDGRSLIIMDLNTNLFNQTI PAAMFQSGKIAANFIAGRRYYVI 600  
 QY 601 KNDGMEKEGCHAGNLEFQGISSEQLNRLSTRNPNCTITSRYVGHSTPTEDNNGSMFLD 660  
 Db 601 KNDGMEKEGCHAGNLEFQGISSEQLNRLSTRNPNCTITSRYVGHSTPTEDNNGSMFLD 660  
 QY 661 MSYNMLSGYIPKEIGSMPTFLINLGHNDISGSIPEVDGKGLNLLDSNKLDERLFO 720  
 Db 661 MSYNMLSGYIPKEIGSMPTFLINLGHNDISGSIPEVDGKGLNLLDSNKLDERLFO 720  
 QY 721 AMSALMTLTETIDISNNNLSGPIPEMGOFETFPAPKEFLANPGLCGYLPKCDPSNADGVAH 780  
 Db 721 AMSALMTLTETIDISNNNLSGPIPEMGOFETFPAPKEFLANPGLCGYLPKCDPSNADGVAH 780  
 QY 781 HOSHGRRPASLAGSVAMGLIFSFVCI FGLILVGRMKRRKKEAELEMYAAGHNSGD 840  
 Db 781 HOSHGRRPASLAGSVAMGLIFSFVCI FGLILVGRMKRRKKEAELEMYAAGHNSGD 840  
 QY 841 RTANNTWKLTGYKEALSTILAAFEKRLKLTADILQATNGFNHNSLISGSGFGDYVA 900  
 Db 841 RTANNTWKLTGYKEALSTILAAFEKRLKLTADILQATNGFNHNSLISGSGFGDYVA 900  
 QY 901 ILKGSAAVAIKGLIHVSQGGREFMAEMETIGIKHNLVPLLGYCKVDEBLLVNEWMK 960  
 Db 901 ILKGSAAVAIKGLIHVSQGGREFMAEMETIGIKHNLVPLLGYCKVDEBLLVNEWMK 960  
 QY 961 YGSLLEDVLODPKKGAVKIKSTRKXIAISARGLAFIHNCSPHIHRDMKSSVLLDEN 1020  
 Db 961 YGSLLEDVLODPKKGAVKIKSTRKXIAISARGLAFIHNCSPHIHRDMKSSVLLDEN 1020  
 QY 1021 LEAVVSDPGMARLMSAMDTHTSVSTLAGTPGYVPPEYVQSFRCSTKGDVYSYGVLLLEL 1080  
 Db 1021 LEAVVSDPGMARLMSAMDTHTSVSTLAGTPGYVPPEYVQSFRCSTKGDVYSYGVLLLEL 1080  
 QY 1081 TGKRPDTPDPEGDNNLVGMVYQHAQLRISDVDPDELMEKDPALIELOHLKVAACLD 1140  
 Db 1081 TGKRPDTPDPEGDNNLVGMVYQHAQLRISDVDPDELMEKDPALIELOHLKVAACLD 1140  
 QY 1141 RAMRPTVQVWAMPEKEIOAGSGIDOSTIRSIEDGFTIEMVMSIKVEPEGL 1196  
 Db 1141 RAMRPTVQVWAMPEKEIOAGSGIDOSTIRSIEDGFTIEMVMSIKVEPEGL 1196  
 RESULT 2  
 P96598  
 Protein P20N2.4 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: P96598  
 R:Phenology: A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso,  
 Chinn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huitart, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Xu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; M01D:21016713; PMID:11130712  
 A:Accession: P96598  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-1166 <STO>  
 A:Cross-references: GB:AB005173; NID:q878502; PIDN:AAE79510.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: P20N2.4  
 A:Map position: 1  
 Query Match 42.4%; Score 2623.5; DB 2; Length 1166;  
 Best Local Similarity 48.5%; Pred. No. 5.4e-112;  
 Matches 573; Conservative 190; Mismatches 341; Indels 77; Gaps 29;  
 QY 33 YREIHOLISFK--DVLDP-KNLLPDM--SSNKNPCTFDEVTRGD-KYTSIDLSKPLNV 86  
 Db 32 FNEVYALLAEKQSVKSDPNNVIGNWKYSGSGSCSWRVSCSDDERITGLDRNSGLG 91  
 QY 87 GFSAVSSSLSLTGLBSLFIENSHIN--GSVSGFKCSASTISLDSRNSLSGPTVLTSL 144  
 Db 92 TLNLV--NLTPALPQLWLYLQGNVFPSSGGDSSSDC--YIQVLDLSNSISDVSMVDYVF 147  
 QY 145 GSCSGLEKLVNVSNTLDFPKVDS--GGIKNSLEVLDLSANSISGANVGVWVLTSDGGELK 203  
 Db 148 SKCSNLYSVNINNNKL--VKGIGFAPESLQSLTVDLSYNIISD-KIPESFISDFPASK 204  
 QY 204 HLAISGNKISGDVDVSRCAVLEFDVSSNNFSTGIPPLGDCSALOHLDSGNLSCGD-FS 262  
 Db 205 YLDLTHNNLSGD-----FSDLS-----FGICMLTFPSLSQNNLSGDKFP 244  
 QY 263 RAISTCELEKLNISNQFVGPFP-----LPKSLQYLSAENKFTGEIPDFLSGACTYL 318  
 Db 245 ITLPNCKFLETINISRNNAKIPNGEYWGSPQNLKQSLAHRNLSGEIPPELSLCKTL 304  
 QY 319 TGLDSGNHFYGAVPFPFGSCSLLESIALSSNPFSGELPMDTLLKMGKLVLDSENEFS 378  
 Db 305 VILDLSGNTFSGELPSQFTACVWLQNLGNVYLSGDFLTVSKITGITLYLYVAANNIS 364  
 QY 379 GELPSLTNLSASLITLDSNNSFPILPVLCO--NPENLTLOEYLONNGFTKIPPTLS 437  
 Db 365 GSVPISTNCS--NLKVLDSNSGFTGNVPSGFCLOSPTVEKLLINNVLSGVVMEIG 423  
 QY 438 NCSELYSLHSFNYSSTIPSSLSGLSKLRDLKLMNMBEGEIPQELMYK--TLETLIL 495  
 Db 424 KCKSLKTIIDLSFNLGTPIPKIIMWLPMVLSLVMMANNLTGTP--EGCVVAGNMLETLIL 482  
 QY 496 DFNLDLGEIPSGLSNCTNLMWISLNNRKTGEIPKMTGRLENLAIKLSNNSSEGNTPDE 555  
 Db 483 NNMLLTGSIPEISISCTYMWIWSLSSNRLTGKIPSGIENLSKALILQGNNSLSGNVPRQ 542  
 QY 556 LGDCSLIWLDTNMLFNQTI PAAMFQSGKIAANFIAGRRYYVIKNDGMEKEGCHAGNL 615  
 Db 543 LGNCKSLIWLDTNLSNNLTGDLPGELASQAGLVMPGSVSGQPAFVRNEG--GTDCRGAAGL 601  
 QY 616 LRFQGISSEQLNRLSTRNPNCTITSRYVGHSTPTEDNNGSMFLDMSYNMLSGYIPKEIG 675  
 Db 602 VEREGIRARLERLPMVHSCPAT--RIYSGMTWYTFPSANGSMYFIDISNAVSGFTIPGIG 660  
 QY 676 SMPYLEFLINLGHNDISGSIPEVDGKGLNLLDSNKLDERLFOAMALMTLTETIDLSN 735  
 Db 661 NMGYLOVNLGNHRLTGTLPDSFGKALGVLDLSHNNLOGLTLPSSLSGLSLDLDVSN 720  
 QY 736 NNLSGPIPEMGOFETFPAPKEFLANPGLCGYLPKCDPSNADGVYAHOSHGRRPASLAGS 795  
 Db 721 NMLTGPITFGGQLTTPPVARYANNISGLCGVPLRPC--GSAPRRPTISRTHAK--QTVATA 777  
 QY 796 VAMGLLPSFVCITGLILVGRMKRRKKEAELEMYAAGHNSGDPTANTMKTGLGYE 855  
 Db 778 VIAGIASFMCFLWALRYVRK--VQKQKQKKEKIEELPTSG-----SCWKLSSVBE 831  
 QY 856 ALSTINLAEPKRLKLTADILQATNGFNHNSLISGSGFGDYVAKILKGSAAVAIKKLII 915  
 Db 832 PLSTINVAETKPPKRLTFAHLAETNGFSAFTMVGSGGGEVYKQALRQGSVVAIKKLIR 891  
 QY 916 VSGGGRFMAEMETIGIKHNLVPLLGYCKVDEBLLVNEWMKGSLEVDYQD--PKK 973  
 Db 892 ITGGGRFMAEMETIGIKHNLVPLLGYCKVDEBLLVNEWMKGSLETVLHEKSSK 951



QY 974 GGVTLKSTRKIAIGARGIAFLHNHNCSPHIIHRDKSSNVLLDENTLBARVDFPGMARL 1033  
 Db 952 GGIVLMAARAKKIIAGARGIAFLHSHCIPIIHRDKSSNVLLDDEDFEARVDFPGMARL 1011  
 QY 1034 MSAMDHLVSTLGTAGYVPEYVDFRCSTGYDVSIVVLLLELLTGKRPDPSDFG- 1092  
 Db 1012 VSADLHLSVSTLGTAGYVPEYVDFRCSTGYDVSIVVLLLELLTGKRPDPSDFG- 1071  
 QY 1093 DNNLVGVKQ-HAKLRISDFVDEPELMKEDPALIEILLKLVAVACLDPRARPRPMQV 1151  
 Db 1072 DNNLVGVAKQLYREKRGAEILDELY-TDKSGVLEFHYIKASQCLDRPRFRPMQV 1130  
 QY 1152 MAMFKEIOAGSGIDSOSTISIEGCGSTIEMVDMKICEVP 1192  
 Db 1131 MAMFKEIMADTEDE-----SLDEPFLKERTP 1156

## RESULT 3

B84431  
 C:Accession: B84431  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: B84431  
 R./lin, X./i, Kaul, S./i, Rounstey, S.D./i, Shea, T.P./i, Benito, M.I./i, Town, C.D./i, Fujii, C.Y./i, M./i, Koo, H./i, Moffat, K.S./i, Cronin, L.A./i, Shen, M./i, Vanaken, S.E./i, Umayam, L./i, Tallon, L./i, Natus, D./i, Nierman, W.C./i, White, O./i, Eisen, J.A./i, Salzberg, S.L./i, Fraser, C.M./i, Venter, J. Nature 402, 761-768, 1999  
 A./Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A./Reference number: A84420; MUID:20083487; PMID:10617197  
 A./Accession: B84431  
 A./Status: preliminary  
 A./Molecule type: DNA  
 A./Residues: 1-1143 <SNO>  
 A./Cross-references: GB:AE002093; NID:g4406778; PIDN:AMD2008.1; GSPDB:GN00139  
 C./Genetics:  
 A./Gene: At2g01950  
 A./Map position: 2

Query Match 40.6%; Score 2509.5; DB 2; Length 1143;  
 Best Local Similarity 46.2%; Pred. No. 8.1e-107;  
 Matches 556; Conservative 183; Mismatches 353; Indels 111; Gaps 25;

QY 9 LSVTLTFFPSFSLSFQASQSLYREIHQIISFQVLPD-KULLPDMGSKRPCTFDG 66  
 Db 14 IQISFIFLTLHSGS-SSSDQSLKTDLSLSFETKMQDDPNILSNMSPKSPCCQFSG 72  
 QY 67 VTCRDQKTSIDLSKPLN-VGSAVSS-SLSLTGLSEFLSNHSHVSGFKC 120  
 Db 73 VTCLOGRVEIMTSGSGSLGIVSFNAFTSLDSLVTKLS-ENFVFN-----STSLLL 125  
 QY 121 SASLTSLSLDRNSLGGPVTTTSLGSCGLKFLVSSNTL-DPQKVGGLKLS-L 175  
 Db 126 PTLTHLEHSSSGLG-----TLPENFSSKYSNLTSTSYNNFTGKLPNDLFLSKKL 179  
 QY 176 EYLDISANSISGANYGVVSLDGGELKHLAISGNKISGVDVGRCVNLEFLVSSNNTS 235  
 Db 180 QTLDSYNNITG-----PISGLTLP----- 199  
 QY 236 TGIPPLGDCSLAHLHDSGNKLSGDFSAISTCTELKLNISNQVGPPIPLP-PLKSL 293  
 Db 200 -----LSSCVMTYLDGFGNSISYIDSLDCTNKLMLSYNNFQGLPKFSGELKLL 254  
 QY 294 QYLSLAENKFFGEIPDFLSGACDTLTGLDSGNHFYGAVPPFGGSLSESLALSNNNS 353  
 Db 255 QSLDLSHRLTGLWIPPEIGDTCRSLQNLKRLSYNNFTVPESSCSGWLQSLDLSNNNS 314  
 QY 354 GELPMDTLLKRGKLVLDLGFNEFSGELPESLTNLSA-SLTLTDLSSNFGSPLIPNL 411  
 Db 315 GPPFVTLIRSGSQTILLNSNLTSGDFP-TSISACKSLRIADFSNRFGSVIIPDLC 371  
 QY 412 QNPKNVTLQELVQNGGFTGKIPPLTSLNCSLVLHLSFNVLSTSTSSGSLKLDLKL 471

Db 372 PGAA-SLEELRLDNLVTGEIPPAISQCESELRTIDLSLNYLNTGTPPEIGNLQKLEOFA 430  
 QY 472 WLNMLEEIPQELMYVKTETLLIDPNDLTGEIPSGLSCTNLMNISNNRLTGEIRKM 531  
 Db 431 WYNNIAGEIPPEIGKQNLKDLILNNQTLGEPPEFPCNSIEWVSFTSNRLTEVRKD 490  
 QY 532 IGRLENALIKSNNPSGNIPDELDCSLTMDLNTLNGTIPPAAMFQSG-KIAAN 590  
 Db 491 FGLISRLAVQLGNNTFTGIBPELQCTTLVMDLNTLHLLGEPPLGQPSKALSG 550  
 QY 591 FIAGKRYVYIKDGMKECHAGNLEFQIGRSBQNLSTNPNITSRVYGGHTSTPE 650  
 Db 551 LLSGNTMAFVAVNG-NSCKGVGLVEFGIRPEKLLQIPSLKCDFT-RMYSGTISLF 607  
 QY 651 DNNGSMFIDMSYNNLSGYIPKEIGSMPTLNLGNDHDSGISIDVEYDGRGNILDL 710  
 Db 608 TRYQTEYLDLSYNQLRKIPDEIGEMIALQVLELSHNQSGEIPFTIGQKNGVDPAS 667  
 QY 711 SNKLDRIPOAMSLTMTLEIDLSNNNTSGPIPEMGQETEPAPKFLANPGLCGPLPRC 770  
 Db 668 DNRLOQITPESFNSLFLVQIDLSNNELTGPLPQSGQSLTPAQVANNPGLCGVPLPEC 727  
 QY 771 DPGNAD--GYAHQRS-HGRPPASLAGSVAMGLFSPVCIFFGLILVGRMRKRRKKEA 826  
 Db 728 KNGNNQLPAGTEEGRAKGRPAASMANSVLGVILISAASCIILVAVARRRRDAD 787  
 QY 827 ELEMYAEHGNSGDGTANNYKLTGYEALSLNLAPEKPLRLTPADLLQATNGFHD 886  
 Db 788 AKMLHSLQAVNSA-----TWKIEKEKEPISINAVTQRLKFSQILKATNGFSAA 841  
 QY 887 SLISGCGFDGVYKALIKDGSAAVKLIHVSQGDREPMAMETIGIKIRNVPLGLYC 946  
 Db 842 SWIGHGGEVPAKATLKQSSVALIKLIRLSCQDRFEMEMETLGKIKRNVPLGLYC 901  
 QY 947 KYGDELLVNEVMKYSLEEDVLDOPKGVK-ILKSTRKRIAIGARGIAFLHNHNCSPH 1004  
 Db 902 KIGBERLLVYEMOYSGLEEVLAGHPPTEGKRRILGWBERKIKKGAAGKGLFHNHCIPH 961  
 QY 1005 IHRDMKSSNVLLDENTLBARVDFPGMARLMSAMDTHLSVSTLGTAGYVPEYVDFRC 1064  
 Db 962 IHRDMKSSNVLLDQDEARVSDGMARLISALDTHLSVSTLGTAGYVPEYVDFRC 1021  
 QY 1065 TKGDVYSYGVVLELLTGKRPDPSDFGDNLLVGVYKQHKL-RISVPPBELMKEDPAL 1123  
 Db 1022 AKGDVYSYGVVLELLTGKRPDPSDFGDNLLVGVYKQHKL-RISVPPBELMKEDPAL 1081  
 QY 1124 EI-----ELLQHLKVAACLDPRARPRPMQVAMFKEIOAGSGIDSOSTI 1170  
 Db 1082 SLNEKEGEGGVIVKEMIRYLETALRCVDDPFSKRPMLQVAVSLRELR-GSENNSHSHS 1140  
 QY 1171 RSI 1173  
 Db 1141 NSL 1143

## RESULT 4

148499  
 receptor-like protein kinase-like protein - *Arabidopsis thaliana*  
 N/Alternate names: protein T28U14.220  
 C/Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C/Accession: T48499  
 R./Bevan, M./i, Murphy, G./i, Ridley, P./i, Hudson, S./i, Bancroft, I./i, Mewes, H.W./i, Rudd, S./i, Len  
 submitted to the Protein Sequence Database, April 2000  
 A./Reference number: Z24493  
 A./Accession: T48499  
 A./Status: preliminary  
 A./Molecule type: DNA  
 A./Residues: 1-1192 <BEV>  
 A./Cross-references: EMBL:AL163652  
 A./Experimental source: cultivar Columbia; BAC clone T28U14  
 C./Genetics:  
 A./Map position: 5

A&gt;Note: T28J14.220

Query Match 26.0%; Score 1608.5; DB 2; Length 1192;

Best Local Similarity 35.2%; Pred. No. 8.5e-66; Mismatches 421; Indels 225; Gaps 36;

Matches 453; Conservative 177; Mismatches 421; Indels 225; Gaps 36;

11 VTTLPFFSFFSLSFQASPSQSLVREIHQISFQDVLDPKULLPDW--SSNNKPCCTFDGVT 68  
 4 LTMFLFLFF--SFSSALVDLSSETTSLISFRSLNLSNNVSSASHCDWVGVT 61  
 69 CRDDKYSIDLSKPLNVGSVAVSSLSITGLLESFLNSHNGVS---GFCASL 124  
 62 CLIGRVNLSLPLSLIR--GQIPKEISLKNRLCLAGNPGSKIPPEIMNK---HL 115  
 125 TSLDLSRNSLSGVVTLTSLGSCGKFLVNSNTLDPFGKV--SGGLKLNLSLEVLDSA 182  
 116 QTLDDLSGNSLTGLLPL--LSELPQLYLDSLNN--HFGSLPSPFFISLPLSLSDVSN 171  
 183 NSISGANVGVWVLDGCGELKHA--ISGNKISGDV----- 216  
 172 NSLSGE-----IPPEIGKLSNLSNLMGLNSFGQIPSEIGNISLKNPAAPSCFNGP 225  
 217 ---DVRVCNLEFIDVSNNSFSTGI-----PFYDCCSALQ 248  
 226 LPKXISLKLKHLAKLDSYNPLKCSIPKSFGBLHNLSTLNVSAELLGLIPPELGNCKSLK 285  
 249 HLDISGNKLSGDFRAISTCTELKLNLS--NQFVGPPIPL--PLKSIQYLSLAENKFT 304  
 286 SLMSFNSLSGRLPLELS--EIPILFSAERNLGSGLSPWQKMKVLDLILLANRFS 342  
 305 GEIP-----DFLSGA-----CDITGLDLSGNHHYGVVPPFGSCSL 341  
 343 GEIPHEIEDCPMLKHLASLNSLSISIPRELCSGSLEAIDLNSLISGITEVFDCCSS 402  
 342 LBSLALSNNSGELPMOTLLKMGKLVLDLSPNEFSGELPEST--TNL----- 388  
 403 LGBLLTNQNGISIPED-LMKL-PLMALDLSNNFTGEIPKLSMKSTNLMETFASYNRL 460  
 389 ---SASLITLDSNNNSGPIPLPVLCONPKNTLOELYLQNNNGPFGKIPPTLSN 438  
 461 EGYLPABIGNANSLKRLVLSNQLGGEIPREI--GKLTSLVLMNANMFOGKIPVELGD 518  
 439 CSELVSLHSPVLSGTSSLSGLSKLRDLKLMNMEGEIPQ-----ELMY 486  
 519 CTSLLTLDLGSNNLQGOIPDKITLALOQLCVLSTNNLSGSLPSKPSAYRHQIEMPLSF 578  
 487 VTTLETLIDFNDLTGEIPSGISNCTNLMWLSLNNRLTGEIPKWIIGLENLAILKLSNN 546  
 579 LQHHGIPDLSTYRSLSGPIPELGEGLVVEISLNNHLSGEIPASLSRLTTLTLLDSGN 638  
 547 SFSGNIPDELGDCLSLIWLDTNLTNLTGTTIPAMFKQSGKLAANFIACKRYVYIKNDGMK 606  
 639 ALTGSIPEKMGNSLKLQGLNLTANNQNGHIP----- 669  
 607 KCHGAGNLEEGQARSBEOLNRLSTRNPNCTISRYGHTSPTFNNNGSMFELMSYML 666  
 670 -----BSFGLIGSLVKTNLTKNKLDEVPASLGNLKEITMDLSFNNL 712  
 667 SGYIPKEIGSMPLYFLINLGHNDISGIPDEVGLRGNTLDSNKLDRIPQAMSLYT 726  
 713 SGELSELSTWELVGLYIEONKFTGEIPSELGNLTGLELTVLDSNNLSGELPTKICGLP 772  
 727 MLTEIDLNNNSLSGPIPEMGQFETPPAPFLANPGICGPRLPDROPSNADYGAHHORSHG 786  
 773 NIEFLMLAKNNLRGEVPSDGVCDPSKALLSGKELCG---RVVGSDC-----KIEG 821  
 787 RRPASLAGVAVMGLFS---FVCI FGL--ILVGERMKR--RRKKEALEMYAEG--HG 836  
 822 TKRSAMGLAGMLGFIITVVFSLRRAMTKRVKQDDPERMESRLKGVVDNLYF 881  
 837 NSGDRANTNMWKLGVKEALSLNLAFFKPLKLTADILLQATNGFNHDSLIGSGRSD 896  
 882 LSGSRS-----REPLSINIAMFEQPLKVLKGLDIVEATDHFSSKNITIGDGFET 930

QY 897 VYKAILKDSANVAIKTLINVSOGDSEEFMAEMETIGKIKHRMLVPLLAGYCKVGBERLLVN 956  
 DB 931 VYKACLPGEKTVAVKLSSEAKTOGNREPMEMETIGKIKVGRPLVSLTGYCSSEELLYV 990  
 QY 957 EWMKYSLEEDVLQDPKKGVKTLSTRKIAIGSARGALFELHNSCPHIIHRDMKSNVL 1016  
 DB 991 EYVWNSLDHMLNQGMLVLDWMSKRLKIAVGARGALFLHNGFIPHIHRDIKASNTL 1050  
 QY 1017 LDENLEARVSDFGMARLMSAMDTHTLSVSTLACRGVVPPEYQSPFCSTKGDVSYGVVL 1076  
 DB 1051 LDGDFEPRKVDLFGIARLISACESHVS-TVLAGTFGYIPPEYQSAATTKGDVSYGVVL 1109  
 QY 1077 LELLTGRPDSPDPCGN--NLVWYKQHAKE-RISDVPPELMKEDPALTEILLQHLK 1132  
 DB 1110 LELVTGKEPT-GPFPKESBEGNVLVMAIQKINGKQAVDIDPLTV--SVALKNSQLRLIQ 1166  
 QY 1133 VAVACLDDRAWRRPTWQVYAMFKEI 1158  
 DB 1167 IAMCLAEPTPAKRPNNMLDVALKALKEI 1192

## RESULT 5

T05332

hypochemical protein F18F4.240 - Arabidopsis thaliana

N.Alternate names: hypochemical protein F1C12.60

C.Species: Arabidopsis thaliana (mouse-ear cress)

C.Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 31-Jan-2000

C.Accession: T05332; T04898

R.Bevan, M.; Terry, N.; Ardiles, W.; Buysaet, C.; Dasseville, R.; De Clerck, R.; De i

eues, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, April 1998

A.Reference number: 215408

A.Accession: T05332

A.Molecule type: DNA

A.Residues: 1-1232 &lt;BEV&gt;

A.Cross-References: EMBL:AL022224

A.Experimental source: cultivated Columbia; BAC clone F1C12

R.Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.

submitted to the Protein Sequence Database, February 1998

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-References: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Genetics:

A.Map position: 4

A.Intons: 863/1; 1116/1

A&gt;Note: F1C12.60; F18F4.240

C/Superfamily: protein kinase Ka21; leucine-rich alpha-2-glycoprotein repeat homology; pr

Query Match 23.0%; Score 1419; DB 2; Length 1232;

Best Local Similarity 31.2%; Pred. No. 3.7e-57; Mismatches 441; Indels 224; Gaps 41;

Matches 402; Conservative 221; Mismatches 441; Indels 224; Gaps 41;

QY 9 LSVTTLFFSFFSLSFQASPSQSLVREIHQISFQDVL-----PDKNLDPWSS--NKGPCT 63  
 DB 4 LVLLLFILCFSGD---GQPG--IINNLDQTLRVKSLVTNPQEDDPLRQMSNINYS 59  
 QY 64 PDGYTC-----RDDKYSIDLSKPLNVGSVAVSSSL 96  
 DB 60 WGTGCTNLTGLFVIALNTGLGLTSGISPFWRFPNNLHLDLSSNNL-VG--PIPTALS 116  
 QY 97 SLTGLSLFTSNHNGSVSG-----FKKS 121  
 DB 117 NLTLSLFLFSNQLGEIPSGSLVINRSLRGNELVGDIPTGLVNLQMLALAS 176  
 QY 122 ASLT-----SLDLSRNSLSGVVTLTSLGSCGKFLVNSNTLDPFGKYS 167  
 DB 177 CRTLGPISQGLVVRVQSLIDQNYLEGPIP--AEIGNCSDLTVPTAAENMLN--GTIP 232  
 QY 168 GGL-KANSLEVLDSANSISGANVGVWVLDGCGE--LKHAILSGNKISGDV--DVSRC 221

Db 233 ABLGRLENIETIANNSUTGE-----IPQGLMSQLOYLISLMAOQLOLIPKSIADL 286  
 Qy 222 VNLFEPLDVSSNNFSTGIP-FLDCCSALOHLISGNKLSGDSRAI-STCEELKILINISSN 279  
 Db 287 GNLQTLDSANNLTLGEIPEEFNNNSQLDLVLANNHSGSLPKSICNNNTNLREOLVISTGT 346  
 Qy 280 QPVGPIPP-LPLKSLQYLISLAENKFTGEIPDPLSGACDTLVTGLDSCGNHFGYAVPPFG 337  
 Db 347 QLSGELPVELSKQSLKQDLDSNNSLSGSIPEALFEIYE-LTDLYLHNNTLBEGLSPSIS 405  
 Qy 338 SCSLESLALSSNNFSGELPMDTLTKKRGKLVLDLSENNSEGEIPESLITLSLITLDI 397  
 Db 406 NLTLNQLVLVHNHLEGLPKG-ISAIRKLEVLPLYNRSEGEIPOEIGNCT-SLKNITDM 463  
 Qy 338 SSNNFSGELPMLCONPKNTLOELYLONNGFTGKIPTLNSCELSVLSHFSYLSGTIP 457  
 Db 464 FGNHFEGEIPPSI--GRUKELNLTLLHQLNELVGLPASILGNCHQNLILDLADNQLSGSIP 521  
 Qy 458 SLSGLSLKRLDKLMLMLBEIPEIPELMYKTELTILDPNDLTG----- 502  
 Db 522 SSFGLKLEBOLMLYNSLQGNLPDSLISLRNTRNLSHNRNGTTHPLCGSSVLSYSD 581  
 Qy 503 -----EIPSGLSNCTNLNWLISLNNRLTGEIPEKIGLEMLATIKSNNFSGNTIPD 554  
 Db 582 VTNNGFEDIEIPLEGNSQNDRLRLGKQULTKIPWTLKIRLSLIDSSNALTGTIPL 641  
 Qy 555 ELGDCRLLMLDLNTNLPNGTIPAMFK--QSGKLIANFLAGRRYYIKNQDKKECHGA 612  
 Db 642 QLVCKKLTHIDLNNNFLSGIPEPWLGKLSQLELK--LSSNQFV--BSLPELPLNC 694  
 Qy 613 GNLL-----EFQGRSEOLNLTSTRNPNITSRYYGHTSPPTDNNSSMMFLDMSTYM 665  
 Db 695 TKLLVLSDGNSLNGSLPOEIGNLGALNLTNDKQFSSLPQAMKLSKLELRLSNN 754  
 Qy 666 LSGYIPEIGSMPL-PIMLGNHDSIGSIPDEVDLRLGNTILDSNNKLDGRIPQAMSA 724  
 Db 755 LTGEIPEIQLQDLQSLDLSYNNFTGDIPTIGLSTLETLDSHNOLGSEVGSIGD 814  
 Qy 725 LTMLETLDSNNLSPPIEMOGFETFPRAKTLNPGCLGYPPLPCDPSNADGVAHORS 784  
 Db 815 MSLGVLNVSFNNLGGKLRK--QFSRMPADSFLLGNTGCGSPLSHCN----- 859  
 Qy 785 HGRPPASLAGSVAMGLFSEVFCIFLIVGRMRKRR--KKEALEMYAEHGNSGDR 841  
 Db 860 ---RVRTISALRALGIMLIVLF-----FKQHDFFK-----VHGSTAYT 899  
 Qy 842 TANNTWKLTGYKALSLNLAPEKPLRK-----LTFADLLQATNGFNHDSILSGGF 894  
 Db 900 SSSSSS-----QATHKPLFRNGASKSDIMMEDIMETNHLSEBPMISGGS 945  
 Qy 895 GDVYKAILKDSAVAIIKLLIHSQ--QGDREPAEMETIGKIKHNLVPLIGCYGVGE-- 951  
 Db 946 GAVYRAELENGTAVAKKLMKODIMSNKNSFSEVYKTLGRIRHHLVLMGVCSSKEGL 1005  
 Qy 952 RLIVNEVWYGSLEVDLPK-----KGVKLTSTRKIAIGSARGLAFLHNSSPHIT 1007  
 Db 1006 NLLIYEYMGNGSIPWMLHEDKRVLEKKKLLDMWAKRLIANGLAQVBYLHHDVPIIVH 1065  
 Qy 1008 RDKSSNVLTDENLEARSVDGEMALMSA-MDTHLSVT-LAGTPGVVPEYQSPFCST 1065  
 Db 1066 RDKSSNVLDSNNMFAHGLDGLKAVLTENCDDTWTDSWTFACGYIABEYVSLKATE 1125  
 Qy 1066 KGDVYSYGVLLLELTGKRPDPSDFGDNMLVGVWKOAKLRIS--DVVDPELMKEDPA 1122  
 Db 1126 KSDVYSWGIIVMEITVGMPDTSVFGAEMQVWVETHLEVASARDKLDLPKLP 1185  
 Qy 1123 LELELQHLKVAVACLDRAMRPTMVO 1150  
 Db 1186 EBDACQVLEIALQCTKTSPOERPSRQ 1213

RESULT 6

C96772  
 Probable receptor protein kinase FIM20.4 [Imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C/Accession: C96772  
 R/Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000  
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.;  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, U.S.; Maiti, R.; Marzalli,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Saitberg, S.L.; Schwartz, J.R.; Shum, P.; Southwick, A.M.; Sun, H.; Tallon, I.  
 ker, M.; Wu, D.; Yu, G.; Frazer, C.W.; Venter, J.C.; Davis, R.W.  
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: A86141; MUID:21016719; PMID:11130712  
 A/Accession: C96772  
 A/Status: Preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1079 <SNO>  
 A/Cross-references: GB:AB005173; NID:96539238; PIDN:AAF15908.1; GSPDB:GN00141  
 C/Genetics:  
 A/Map position: 1  
 Query Match 22.3%; Score 1380; DB 2; Length 1079;  
 Best Local Similarity 32.2%; Pred. No. 1,96-55;  
 Matches 392; Conservative 204; Mismatches 406; Indels 214; Gaps 44;  
 Qy 26 ASPQSLYREIHQLISFKVLPDPK-----LLPQSSKNP--CTPQVTC--RDKYTS 76  
 Db 5 AVAGDSDLDREVLSTLSYLSRNPQKGLYTKEMENQVVCQWPEITCTPPRSRYTG 64  
 Qy 77 IDLS-----SKPLNVSFVSSLSLTGLSLFLSNHINGSVSGFKSALITLDSRN 132  
 Db 65 INLTSTISGPKFKNSALT-----ELTYLDSRN 94  
 Qy 133 SLSPYTLTISGSGSGKFLNVSNTLD-----PPKYSGLKLNLSLEVDLSANISGA 188  
 Db 95 TIEGEIP--DLSRCHNLKHLNHLNLEGLSLPG-----LSNLEVDLSINRITGD 145  
 Qy 189 NVGVWVSDGCGELKHLAISGNKISGVD--VSRVNLFLVDVSSNNF-----TGI----- 238  
 Db 146 IOSSPPLF--CNSLVANLSTNNFTGRIDIDFNGCRNLKTYVDSSNRFSEWVTGGRLV 203  
 Qy 239 -----PFLGCSALQHLDISGNKLSGDFPSRAISTCTELKLNISNQFV 282  
 Db 204 EFSVADNHLSGNISAMFRNGCT-LQMLDLSGNAFGGEFPGOVSNCONVNLMLQNKFT 262  
 Qy 283 GPIPL-PLKSLQYLSLAENKFTGEIPDFLSGACDTLVTGLDSCGNHFGYAVPPFGSGS 340  
 Db 263 GNIPEIGSISLKGIVYGNTPRDIPTILLNLTN-LVFLDLSRKKFGGDIQEIFGRFT 321  
 Qy 341 LLESIALSSNNFSGELPMDTLTKKRGKLVLDLSENNSEGEIPESLITLSLITLDI 400  
 Db 322 QVKYLVHANSYVGVGINSNLIKLPNLSRLDLDGNNPFGQLPFEISQI----- 369  
 Qy 401 NFSGELPMLCONPKNTLOELYLONNGFTGKIPTLNSCELSVLSHFSYLSGTIPSSL 460  
 Db 370 -----QSLKFLILANVNFSGDIPQEGNMPGLOALDLSFNKLTGSIPIASF 414  
 Qy 461 GSLKRLDKLMLMLBEIPEIPELMYKTELTILDPNDLTGIPSGLSNNTLWISIS 520  
 Db 415 GKLTSL--LM-----LMLANNLSGEIPEIENGCTSLMFWNA 450  
 Qy 521 NNRLTGEIPKIGRL-ENLAILKLSNNFSGNIPDELGDCSL-TWLDLNTNLPNGTIPA 578  
 Db 451 NNQSLGFFHELTRMGSNPSPTEVVRQNKDKLIAGSGCLAMKWIIPAEFPFR-FYVA 509  
 Qy 579 AMFQSGKLIANFLAGRRYYIKNQDKKECHGAGNLLEFQIRSEQLNRLSTRPCNIT 638  
 Db 510 ILTKGCRSLMDHY-----LKGYGLFPVC-SAGST-----VRTLKISAY-----QLS 551



QY 1135 VACLDRAWRPRTWYVAMFKEI 1158  
Db 985 CLCISENKKQRPQTQOLVSWLDDV 1008

## RESULT 8

E86308  
hypochemical protein F20D23.7 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_rev1510 02-Mar-2001 #text\_change 31-Dec-2001  
R/Accession: E86308  
A/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Hughes, B.; Hiltz, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzalli, R.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Roederberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Accession: A86141; MID:21016719; PMID:11130712  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1133 <STO>  
A/Cross-references: GB:A8005172; NID:95734762; PIDN:AD50027.1; GSPDB:GN00141  
A/Map position: 1

Query Match 22.1%; Score 1367; DB 2; Length 1133;  
Best Local Similarity 32.9%; Pred. No. 7, 8e-55;

Matches 398; Conservative 179; Mismatches 455; Indels 176; Gaps 36;

QY 8 FLATVILFFSFFSLSFQASPSQSLYREIHQIISFKDVLDPDKL LPPDMS-NKXNCTFD 65  
Db 7 FLATVILFFSFFSLSFQASPSQSLYREIHQIISFKDVLDPDKL LPPDMS-NKXNCTFD 65  
QY 66 GVTYCRDCK-VTISIDSKPLVGFSAVSSLSLTGLSLFLSNHNGSV-SGPKCSAS 123  
Db 60 GIACHTLKRVTVDLNGMNL---GTLSPICLCKLHKLKLVSTNFGPPODLSLCS 116  
QY 124 LVSIDLSNLSG-----PVYTLTSL-----GSCGSLKPLVANSNTLD 161  
Db 117 LEVLDLCTRRFHVGIPIQLTMTITLKKLYLCENYLFSGIPROIGNLSLQGLVYSNNL 175  
QY 162 PRGKVSQGL-KXNSLEVLDLSNLSGANVYGVWVSDCGGELKHLAISGNKISGV--DV 218  
Db 176 -GVIPPSMAKLRQRLIRPAGRNFGS--VIPSEIS-GCESIKVTLGLAENLLEGSILPKQL 231  
QY 219 SRCVNLLEFLDVSSNNFSTGI-PFLGDCALQHLIDISGNKLSGDSRAISCTELKILNIS 277  
Db 232 EKLQNLTLQILMQLNQLSBEIPSVANIRLEVLALHENVYFGSIPRIGKLTWKRLVLY 291  
QY 278 SNQFQFIPPL-PLKSIQYISLAENKFTGIP-DPLSGACDTLTGLDLSGNHYGAVP 334  
Db 232 TNLQGLIPREIGNLIDAEIDFENQLTGFI PKEF--GHILMKLHLLENLILGPIR 349  
QY 335 FFGSCSLLESIALSSNNSGELPMDTLLKXGKLAVDLSPNBSGELPESLTVLSASLIT 394  
Db 350 ELGETTLLEKDLISNRLNGITPQ--LQFLPYLDLQFLNQLGKLP-PLIGFYNFSV 407  
QY 395 LIDSSNNSFGILPNTLQNPNTLQELVYLQNGFTGKIPTLNSCSELVSLHSFYVLSG 454  
Db 408 LDMNSNISGPIPAHFCR--FQTLILSLSNKLSGNIIPDLDTCKSLTGLMAGNQLTG 465  
QY 455 TTPSLSGLSLRLDKLMLNLEBEIPOELMYVKTETLLDNLDTLGEIPSGLSNCTNL 514  
Db 466 SLPTLEFLQNLALHQLNQLSISADLGLKLNLERLANNFNGEIPRIGNLTGI 525  
QY 515 NWISLNNRLTGEIPKVIKGRLENLAIKLSNNSPSGNIPELDCDGLSLWDLNTLWFLNG 574  
Db 526 VGFNLSNQLTGLHFKELSGCVTTIQRDLDSGNKPSGTYIAQELQVLYLEILRLSDRLTG 585

QY 575 TIPAMFKOSKTIANFIAGKRVVYIKNDGKCKGAGNLEFGGISEQINLRISTNP 634  
Db 586 EIP-----HSFGDL-----TR-- 596  
QY 635 CNITSRVYGHTSPTEPDNNSMFMFLDMSYMLSGYIPKEIGSMPLFI-LNLGNHDSGS 693  
Db 597 -----LMEIQGNLSSENIPELGLTSLQISLNSHNLSGT 635  
QY 694 IPPEVGDIGKATLIDSSNKLDRIPQAMSALTMLTEIDLSNNLSGPIPEWQFETPP 753  
Db 636 IPDSLNQLQWLEILYLDNDKLSGEIPASIGNLSMLICNISNNLVGVPTPAFQRNDS 695  
QY 754 AKLNPNGGCVYLPKCDP--SNADGVAMHQRSHRRRPAASA-----GSYAMGLTSFV 805  
Db 696 SNAGNHGICNQSRSHCPVLVHSDSKMLWILNSQROKILITTCIVGVS--FLITFL 752  
QY 806 CIPGLILVGRERKRRRKKKALEMYAEGHNSGDRTPANTNMLKLVKEALSINLAPE 865  
Db 753 GLCWTI-----KREPAFVLE-----DQTKPDV-----MDSY 781  
QY 866 KPLKLTFLADLQATNGFNDLSGSGFDVYKAILKDSAVAIIKKLHVSGG--DR 922  
Db 782 FPKGFTYGLVDATRNPSBDVLGRGACGVYKAEVSGEVIATKLT-NSRGRGASDN 840  
QY 923 EPMAEMETGKIKHNLVPLGCKVGERLTVNEMKYSLEEDVDDPKKGVKLTST 982  
Db 841 SFRALSTLGRHNRNIYKLVGFCRQNSNLLTYEYMSKSLGBOQLRGKNCIL-DWNA 899  
QY 983 RKLIAGSARGLAFLHNCSPIIHRDKSSNNVLDENLEARYSDPFGMARLMSAMDTHLS 1042  
Db 900 RYRIALGAEEGLCYLHDCRPOIVARHDKSNLLIDRRPQHVGDPFLATLID-LSTSKS 958  
QY 1043 VETLAGTGYVPEPYGOSFRCSYGVYVYVLELTGKRPDSEDFPDNNLVGVWVQ 1102  
Db 959 MSAAVAGSTGYLAPAYWVKTEKCDIYSFGVVALLELTITKRPVQPLEQG-GDLVWVVR 1017  
QY 1103 HAKLRIS--DYVDELMKEDPALIEIILQHLKVVACTLDRAWRPRTWYVAMFKEIQ 1160  
Db 1018 SIRNMIPTLMEFDARLDNDKRTVHMSLVKTLRFTCSNPSASRPTREVVAMITEARG 1077  
QY 1161 GSGIDSQS 1168  
Db 1078 SSSLSSSS 1085

## RESULT 9

B84742

probable receptor-like protein kinase [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_rev1510 02-Feb-2001 #text\_change 02-Feb-2001  
C/Accession: B84742  
R/Li, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; F. euse, D.; Niemann, W.C.; Cronin, L.A.; Shen, W.; VanAken, S.E.; Umayam, L.; Tallon, L.; Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Accession: A84420; MID:20083487; PMID:10617197  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1124 <STO>  
A/Cross-references: GB:A8002093; NID:92924777; PIDN:AC04906.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2g31170  
A/Map position: 2

Query Match 22.1%; Score 1364; DB 2; Length 1124;  
Best Local Similarity 33.1%; Pred. No. 1, 1e-54;

Matches 395; Conservative 177; Mismatches 483; Indels 138; Gaps 34;

QY 2 KTFSSFLSTVTLFFSFFSLSFQASPSQSLYREIHQIISFKDVLDPDKL LPPDMS 56

```

Db 10 KESKMEVGV-----LFLTLVWTSLSLSDGQFILELKNRGFODL---NRHNMN 59
Qy 57 S-NKNDCTDGTCTRDDKYTSDLSKPLNNGFSANVSSLSLGTLESFLNSHINSV 115
Db 60 GIDETPCNMIGVNC-----SSQ-----GSSSSNSLV----- 86
Qy 116 SGFKCASLTLSDLSRNSLGGPYTLTSLGSCGLKFLVNSNTL--DPPKVGSGGLKLN 173
Db 87 -----VTSLSLSMNLGIVS--PSIOGLVNLVYMLAVNLTGDI-PREIG--NCS 133
Qy 174 SLEVLDSLANSIGANVGVNLSGCGELKHLAISGNKISGDV--DVSKCVLPELDVSS 231
Db 134 KLEVMFLNNQPGGSIIVE---INKLSQLSFFNLCNNKLSGSLPEBIGLYMLEEVAVT 190
Qy 232 NNSTGIP-FLGDCSALQHLDISGNKLSGDFSAISTCTELKLNLSNQFVPIPL-- 288
Db 191 NMLTGPLPSLGNLNTKLTTRACONDFSGNI-PTEIGKCLNKLGLAQNISELKEIG 250
Qy 289 PLKSLQVLSAENKFTGEIPDPLSGACDTLTGLDLSGNHYGAVPPFGSCSLLESIALS 348
Db 251 MLVLTQVILMOKNFSGTIPKDI-GNLTSETLALYGNLSVGPISPEIGNMKSLKLYLY 309
Qy 349 SNNFSELPMDTLKKGKVLDSFNEFSGELPESTLNLSALTLTDLSSNNFSGPIIP 408
Db 310 QNQLNGTIPKE-LCKLSKWEIDFSENLISGEIPELISKIS-ELRLYLYQNKLTG-IIF 366
Qy 409 NLGNPKMTLOELYLQNNNGFTGKIPPTLSNCELSVLSLFFNYLSGTIPSSLSLSKLD 468
Db 367 NELSKLRN-LAKLULSINSLTGPIPPGQNLTSWRQQLFPHNSLSGVIPQGLGYSPLV 425
Qy 469 LKLMNMLEGIPOLMAYKLTLETLIDFNDLTGEIPSGLSNCTNLMVLSNNRLTGEI 528
Db 426 VDFSNQSGKIPPIPCQSNMILNLGNSRNI-FGNIIPGVLRCKSLQLRVAGRLTGQF 485
Qy 529 PKWIGLENLALIKLSNNSFSGNIPDELGDORSLIMDLNTNLNFGTIPAMFKQSG-- 585
Db 486 PTELCVNLVLSLDELQRFSGSLPEIGTCOKLORHLAANOSSNLPNISKLSNLT 545
Qy 586 -KIANFAGKRYVYIKNDGKKECHGAGNLEFQIRSEQLNRLSTRNPNCTISRYVYG 644
Db 546 FNVSNLSLTGPIPSRIANCKMLQRDLSRN--SFISGLPELGLHOLEIRLSENFRSG 603
Qy 645 HTSPFFDNNGSMFLDMSYNNLSGYIPEKISGMPYLF-I-MLGNHDSGSI-PDEVGDRLG 703
Db 604 NIPFTIGNLTLTELQMGNLPSGIPPOLGLSSLOLAMLSTYDFSGELPPEIGNHL 663
Qy 704 LMIILSSNKLGRIPQAMASALTMLTEIDLSSNNLSGPIPEMGQFETPPKAFNLNPGLC 763
Db 664 LMYLSLNNHLSGELPTTPTENLSLGCNFSYNNLTGQLPHQIPLQNMTLTSLFNGKGLC 723
Qy 764 GYPLPRCDSNADGYAHHRSHGRAP--ASL-AGSVANG---LFSVYCIPLGLVGRE 816
Db 724 GGHLSRCDP-----SHSSWPHISLSLAKGASRRRIIIVSVYIGISILLIIV 772
Qy 817 MKRRRRKKALEMAVEAGHNSGDSRTAANTNMKLTGVKEALSINLAPEKPLKLTPLADL 876
Db 773 VHELRMPVEPTAPYVD-----KAPFQESDIYVPERKTIVDI 812
Qy 877 LQATNGFHNDSLSGSGFGVYKAILKDSAVAIAKL-----IHSVQDGRFEMAEWE 929
Db 813 LEATKGFHDSYVGRACGVYVYKAWPBGKTIYVKLJESNRREGNNNSNNTDNSPRAETL 872
Qy 930 TTGKIKHRLVPLLYC--KVGDELLVNEVMKYSLEVDLPDPKKGAYKALKSTRKKA 987
Db 873 TLCKIRHRNIVRLYSFCYHSGNSNLLIYEMVRSGLIGELHGGKH--SMOMPTRFATA 930
Qy 988 IGSARGIAFLHNSCPHIIHRDMKSSVLLDETLARVSPFGMARLSAMDTHTSVSTLA 1047
Db 931 LGAAGGLAYLHDOCKPRIIHRDKSNMILIDENFEHAVGDFGLAKYID-MPLSKSASAVA 989
Qy 1048 GTFGYVPEPEYOSFRCTKGDVYSYGVVLELLTGKRPDTSDPGDNNTLVGKQYQAK-- 1105
Db 990 GSYGILAPETAYAMKYTEKCDIYSFGVVLLELLTGKRPVQPLRQG--GDLATWTRNHIRDH 1048

```

```

Qy 1106 LRISDVFDEELMK-EDPALEIELLOLKVAVACLDDBAMRPTMVOVMAFKE 1157
Db 1049 SLTSEILDPTLTKEEDVILNMHTVTKIAVLCTKSSPSDRPIMREVVMLIE 1101

RESULT 10
105898
hypoetical protein FEH11.170 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 22-Oct-1999
C/Accession: T05898
R/Beyan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Bancroft, I.; Mewer
submitted to the Protein Sequence Database, April 1998
A/Reference number: 215456
A/Accession: T05898
A/Molecule type: DNA
A/Residues: 1-1003 <BEV>
A/Cross-references: EMBL:AL021684; GSPDB:GNO0063; ATSP:FEH11.170
A/Experimental source: cultivar Columbia; BAC clone FEH11
C/Genetics:
A/Gene: ATSP:FEH11.170
A/Map position: 5
A/Introns: 867/1
C/Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; pI

Query Match 22.0%; Score 1357.5; DB 2; Length 1003;
Best Local Similarity 33.0%; Pred. No. 1.8e-54;
Matches 384; Conservative 177; Mismatches 382; Indels 221; Gaps 34;

Qy 14 IFFSFFSLSPQASPOGLYREIHQISFKVLT---PDKKL-LPDMSSNKNPCTPGVY 68
Db 7 LFLFIHISHTFAPSRPIS---EPRAILSLKTSITGAGDDKNSPLSMKVTSTFTWIGVT 63
Qy 69 CRDDK--VTSIDLSKPLNNGFSANVSSLSLGTLESFLNSHINSVGFKCSASLTS 126
Db 64 CVSRHRVHVSIDLSG-----LNLSTLSVDVSHRL-----LQN 97
Qy 127 LPLSRNSLGGPYTLTSLGSCGLKFLVNSNTLD--PPKVGSGGLKLNLSLEVLDSLANS 184
Db 98 LSLAENLISGPIP--PEISSLSGLRHLNLSNNVFNFGSPDEISGLV--NLRVLVDVYNNN 153
Qy 185 ISGANVGVNLSGCGELKHLAISGNKISGDVAVRCYNLELDVSSNNFSGTIPFLGDC 244
Db 154 LTGDLPLVS---VTNLQTLRHLHLAGNYFAGKIP-----PSYGSW 189
Qy 245 SALQHLDISGNKLSGDFSAISTCTELKLNLS-SNOFGPIPL--PLKSLQVLSLAEN 301
Db 190 PVEYLAVSQNELVVKIPEIENLTLRELTYGYNAFEDGLPPEIGNLSLVRFDGANC 249
Qy 302 KFTGEIPDFLSGACDTLTGLDLSGNHYGAVPPFGSCSLLESIALSSNNFSGELPMDTL 361
Db 250 GLTGEIPPEI-GELQGLDTLFLQVNVFSGPLTWELGTLSLSMDSNNMFTGEIPLA-SF 307
Qy 362 LKRGKVLDSLNERFSGELPESTLNLSALTLTDLSSNNFSGPIPLNCONPKMTLOEL 421
Db 308 AELKNLTLNLPNNKJHGEIPEFIDPL-DELEVYLQWENNFTGSIPOKLGKNGKLTVD- 365
Qy 422 YLQNNNGFTGKIPPTLSNCELSVLSLFFNYLSGTIPSSLSLSKLRDLKLMNMLEGIP 481
Db 366 -LSSNKLTLPLPMMCSGNLTLETLITLGNFLPSGIDSGKCSLRIRIMGENFLNGSIP 424
Qy 482 QELMYKLTLETLIDFNDLTGEIPSGLSNCTNLMVLSNNRLTGEIIPKIGLEMLALI 541
Db 425 KGLFGLFKLTQVELQDNYLSGELPLVAGGVVNLGOISLSSNNQSGPLPPAIGFTGVQKL 484
Qy 542 KLSNNSFSGNIPDELGDORSLIMDLNTNLNFGTIPAMFKQSGKIAAFIAGKRYVYIK 601
Db 485 LLDNKNFGPIPSFVG----- 500
Qy 602 NDGKKECHGAGNLEFQIRSEQLNRLSTRNPNCTISRYVGGHTSPTFDNNGSMFLDM 661
Db 501 -----KLQQLSK-----IDF 510

```



QY 662 SYNNLGVYIPEKIGSMPLFILTANGHNDISGSDPEVDGLGNTLIDLSNNKLDGRIPQA 721  
 Db 511 SHNLSGRIABEISRCKLITFVDSLNRNELSGEIPBEITAMKILYNLSRNHLVGSIPGS 570  
 QY 722 MSALTMLEIDLSNNLSGPIPEMGQFETPPAPKLANPGICGYPPLPCDSNADGIAH 781  
 Db 571 ISSMOSLSLDPSTYNLSGLVPGTQGSFYNTSFLGAPDLCCGYLPGCKDGVAKG-GHQ 629  
 QY 782 QSHGRRPASLAGSVAMGLFSPVCIPLILVGRMKRRKKEALEMYAFEGHNSGDR 841  
 Db 630 SHSKGPIASNMKLLIVLIGL---VCSIAFAVVA-IIRKSLKASE----- 671  
 QY 842 TANTNMKLTGVCALSLNLAFAEKPLKLTFADLLQATNGFNHDSIIGSGGFDYVKAI 901  
 Db 672 ---SRAMKLT-----AFQR-LDFTCDVL---SLKEDNIIIGKAGIYVKGV 742  
 QY 902 LKDGSAVAIKKLIHVS---GGQDREFMAEMETIGIKRNLVPLIGYCKVDERLNVEM 959  
 Db 713 MPNDOLVAVKRLAMSGSSHDHGFNAEIQTLKRIHRHIVRLGFCSEHETNLLVYEM 772  
 QY 960 KYGSLVEDLQDPKKGKYLKLTSTRKIAISGARGLAFLHNHNSPHIIRDMKSSNVLLDE 1019  
 Db 773 PNGSLGEVLHG-KKGG-HLHWDTRYKIALBAKGLCYLHDCSPLIYHRDVKNNILDS 830  
 QY 1020 NLEARVSDFGARLMSAMDHLSTLACTPGYVPEPYQSFRCSCTGYYSYGVILLEL 1079  
 Db 831 NFEHVAHDFGLAKFLQDSGTSECSALAGSYGYIAPAYATLLKVDKESDYVSGVILLET 890  
 QY 1080 LTRKRPDSDPDGDN-NLVGMVKQHA---KLRLSDVDPDELMKEDPALIEILLQHL-KYA 1134  
 Db 891 VTGKRKFGV--EFGDGVDIYQWVRKTDMSNDVYLKVIDPML---SSIPHEVTHAVYA 944  
 QY 1135 VACLDPRAMRPTMVQVAMFKEI 1158  
 Db 945 MLCVEQAVRPTMRREVQILTEI 968

RESULT 11  
 967746  
 hypothetical protein T9N14.20 [imported] - Arabidopsis thaliana  
 C1Species: Arabidopsis thaliana (mouse-ear cress)  
 C1date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C1Accession: G967746  
 R1Theologos, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 Hansen, N.F.; Hughes, B.; Hultzer, L.  
 N1Ref 408, 816-820, 2000  
 A1Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A1Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
 A1Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A1Reference number: A66141; MUID:21016719; PMID:11130712  
 A1Accession: G967746  
 A1Status: preliminary  
 A1Molecule type: DNA  
 A1Residues: 1-1095 <STO>  
 A1Cross-references: GB:AE005173; NID:g10645385; PIDN:AA621504.1; GSPDB:GN00141  
 C1Genetics:  
 A1Gene: T9N14.20  
 A1Map position: 1

Query Match 21.8%; Score 1349.5; DB 2, Length 1095;  
 Best Local Similarity 32.4%; Pred. No. 4.7e-54;  
 Matches 395; Conservative 192; Mismatches 398; Indels 233; Gaps 42;

QY 3 TFSFSLST-----TLFSPFSLSFQASPSQSLVREIHQLSFDVLPDKLL 52  
 Db 35 SLSVFLLTSEAVCNLQDRDSLWMS-GNVSSPVS-----LH----- 71  
 QY 53 PDMSNKNPCTFDTGVTCT---RDDKVTSLDLSKPLNVGSAVSSSLSLTGLSLFLSNS 109

Db 72 ---WNSSIDCCSMEGISCDSPENRVTISILSSRGIS---GNIPSSVLDLQRLSRIDLHN 126  
 QY 110 HINGSV-SGEKCSA-SLISLDSRNSLSGPVTTTLTSLGSCGLKFLNVSNTLDPGKVS 167  
 Db 127 RLSEGLPGLFSLALQDLVLVDLSYNSFKGELPLQOSFGN-----GNSNI-PP----- 172  
 QY 168 GGLKNSLEVLNLSNLSISGAVVGVWLSDCGGLKHLAISGKISGDVDSRCVNLLE 227  
 Db 173 -----IQTVLDSN-----LLEBEIILSSVPLQGANLTSF 203  
 QY 228 DVSSNNFSTGIP-FLDGSA---LOHLDSGNKLSGDFSRASITCTELKLNISNGV 283  
 Db 204 NVSNNSFTGISIFM---CTASPOLTKLDPSTYDFSGDLSGELSRCSRLSVRAGNNLSG 261  
 QY 284 PLPP--LPLKSLQYLSLANKTGEI PDLFGACDTLGLDLSGNHFYGAVPPFGCSL 341  
 Db 262 ELPKEIYMLPELEOLFPLPNRISGKIDNGIT-RLTKLTLEIYSNHIGEIPKDIGKLSK 320  
 QY 342 LESLALSSNNSFSGELPMDTLKMRGLKVLDSFNEFSGELPESLNLNLSLITLDSNN 401  
 Db 321 LSLQLHVNLMGSLPV-SLANCTKLVKLNLRVNLQGLTSLAIDFSRQSLSLDLG--- 376  
 QY 402 FSGPILPMLCONPKNTLQELYLONNGFTGKIPTLSNCSLVSILHSFNYLSGTIPS--- 458  
 Db 377 -----NNSFTGEFPTVYSCKMTFARFAGKLTGQISQVL 413  
 QY 459 SLGSS-----KLRLKMLNLSGEIPEOLMYVKTLETLIDPDLGEIPSG----- 507  
 Db 414 ELESLSFTFSDNKNNTNLGALSLIOG-----CKKLSTLIWAKNFYDETVSNDFL 465  
 QY 508 -LSNCTNLMWISLSNNRLTGEIPKMGLENTLALIKLSNNSFSGNIDELGDCSLTMD 566  
 Db 466 RSDGPPSLQIGIGACRLTGEIPAMLIKQKVEVMDLSMRPFVTIGMGTLPDLFLD 525  
 QY 567 LNTNLFNGTIPAMFQSGKIAANFIAGKRVYIKNDGKKECHAGANLLEF-----Q 619  
 Db 526 LSDNFLTGELEKELFQLRA-----LMSQKAY-----DATER-----NYLELPVAVPN 569  
 QY 620 GIRSEOLNRLSTRNPNCTITSRVYGHSSPTFDNNGSMFLDMSYNNLSGYIPEIGSMY 679  
 Db 570 VTTQOQVNLSSLPTTYIKR-----NNLTGTIPVEVGQLV 606  
 QY 680 LFIINLGHNDISGSDPEVDGLRGLNTLDSNNKLDGRIPQASALTMLEIDLSNNLS 739  
 Db 607 LHLELLGNNSGSSIPBELNLTNLELIDLSNNLSRLIWSLGLHFLSYEVANNTLS 666  
 QY 740 GPPEMGQFETPPPAKFLNPNGLC-GYPLPRCDPSNADGYAHHQSHGRRPASLAGSVAM 798  
 Db 667 GPITGTQFPDTPFANFEGNPLLCGVLTLSQDPTQ---HSTTKMGKKNRTLVIGLV 723  
 QY 799 GLSPFCIFGLILVGMKRRR-----KKEALEMYABGH-----GNSGRTNNTN 847  
 Db 724 GLFPGVLI--LVLLALVLSSKRRVNGDESENELEINSNGSYSEVPAGSDK----- 774  
 QY 848 WKLTGVKALSLNLAFAEKPLKLTFADLLQATNGFNHDSIIGSGGFDYVKAIKDGSA 907  
 Db 775 -----ISLVLLFGNSRRE--VKDITIFELLKADNSQANIIICGGFGLYKATLNGTK 827  
 QY 908 VALKLIHVSQGDREFMAEMETIGIKRNLVPLILGYCKVGERLLVNEVMKYSLEDV 967  
 Db 828 LAVKQLTGDYGMKEKEKAEVETLSRAKHENLVAALQGYCVHDSARILIYSFMENGLDYM 887  
 QY 968 LQDPKKGVYLKLTSTRKIAISGARGLAFLHNHNSPHIIRDMKSSNVLLDENLARYSD 1027  
 Db 888 LHNHNPBQALDMPKLTNLMKAGSSGLATMHCCEPHIYHRIKSNILLDNFAYVAD 947  
 QY 1028 FGMARLMSAMDTLSTLACTPGYVPEPYQSFRCSCTGYYSYGVILLELITGKRPD 1087  
 Db 948 FGLSRLLIPRTHVT-TELVTGLGYTPPEYQAMVATLKGDDYVSGVWMLLEITKRPME 1006  
 QY 1088 -SPDGDNNLVGMVKQHAKR---ISDVPELMKEDPALIEILLQHLKVAVACLDRRA 1142



```

Db      6 SNFF-----FLFPCSVSMAPQPTLSLSDGQALSLK--RSPSPSLFSSWDQDQTPCS 57
Qy      64 FDVTC-RDDKVTISIDSKPLNVFSAVSSSLTGLTSLFLSNHNGSV-SQFKKS 121
Db      58 WYGITCSADNRVIVSIPDFLNL--SSI-PDLSSLSLQFLWLSSTNLSGPIPPSFGTL 114
Qy      122 ASLTLSDRSLSGPTTTLTSLGSCGLKFLVSSNTL--DPGKVSQGLKLNLEVL 179
Db      115 THLLDLSSNLSGPIF--SELGRSLTQPLILNANKLSGISIPQIS--NLFALOYLC 169
Qy      180 LSANSISGA--NVGWVLSDOCCGELKHLAISGKISGDVDSKVCNLEFLDVSNNFSG 237
Db      170 LQDVLNMGSLIPSSFGSLVS-----LQDFLNGN-----TNLGGP 203
Qy      238 IPIF-LGPCSALQHLIDISGNLSDGFSRASTCTELKLNISNQFVPIPP-LPIKS-LQ 294
Db      204 IPALQGLKMLTTLGFAASGLSGSIPSTFGNLVNLQTLALYDFEISGTLPPQGLCSER 263
Qy      295 YLSLAERKFTGEIPDPLSGACDTLTGLDLSGNHFGYAVPPFGSGSLSLSSNNFSG 354
Db      264 NLYIHNKMLTGSIPKEL-GKLQKITSLTLNMGNSLGSVIPPESINCSLIVFVDSANDLTG 322
Qy      355 ELPMDTLLKXRGKVLIDLSPNERSGELPESLTVLSALTLTLDSSNNSGPIPLMLCONP 414
Db      323 DIPGD-LGKLWMLBQQLSDNMTQIIPWELNCS-SLIALQDKKKSLSGSI-PSQIGNL 379
Qy      415 KNTLOELYLONNGFTGKIPTLSNCSSELYSLHSFNYSGLTIPSSIGSLSKLDKMLN 474
Db      380 K-SLQSFPMENISIGTIPSSFNCTDLVALDLNRKLTGRIPBELFSLKSLKLLIGN 438
Qy      475 MLEGEIPELIMYKTLLETILIDFNLTGELPSGLSNCNLMNISNNRNGELIPKMTGR 534
Db      439 SLSSGGLPKSAKQSLVRLRVEHQSLQIPKEIGELQVLVDLYNHRFGSGPYEISN 498
Qy      535 LENTALLKLSNNSFSGNIPDELGDCHSLTMDLNTNLNFTGTPAAMFQSGKIANFIAG 594
Db      499 IYVLELIDVHNNTITDIPQGLNVLNLEQDLNRSNFTGNIPLSF-----544
Qy      595 KRVVYIKDGMKKECHGAGNLEFQGISBQNLNSTRNPNCTISRYVGGHTSPFDNG 654
Db      545 -----GNL-----SYLNKLIINN-----NLITQIIPKSIKMLQ 572
Qy      655 SMMFLDMSYMLSGYIPKEIGSMPLYFI-LNKGNDISGSI-PDEVDGRGNTIIDLSSNK 713
Db      573 KTLTLDLSYNSLSGELPQELQVTLTINLDSVTFTGNIPEFSSDLTQLOSLDLSNS 632
Qy      714 LQGRIPQMSALTMTEIDLNNNLSGPIPEMGOFETPPPAKFLNNGLCGYPLRCDPS 773
Db      633 LHGDI-KVLGSLTSLASINISCNNSFGPISTPFKTIISTTSYLNQNTLC-----H 682
Qy      774 NADGY---AHORSHGR-----PASLASVAMGLLFSVCI-FGLILVGRERKRRK 823
Db      683 SLDGITSCHTQONNGVKSPIKIVALTAVILASTITAILAALL-----LENNHLY 733
Qy      824 KEAELEMTAEGHNGSGDTANTNTWKLTGVKEALSINLAPEKRLKLTADILQANNGP 883
Db      734 KTSQ-----NSSSPETAEDFSYPTWFI-----PQGLGT-VANNYVTSL 772
Qy      884 HNDLSISGGGFDVYKALKKDSAVAIKLIHVSQGDGDR-----EFMAEMETIGIKIR 937
Db      773 TDENYTGCGSGLYKAIIPNGDIVAVAKMKTKDNNEEGESTIDSPABEIQLIGNIRRH 832
Qy      938 NLVPLLYGCKVDERLLVNEVMKYSLEVDYQDPKGVKLUKSTRKAIISARGIATL 997
Db      833 NIVKLLGYSKSVKLLIYNFPNGNQLQLQ-----GNRLDMETRYKAIIGAAGIATL 888
Qy      998 HHNGSPRIIHDMSNVLDENLEAVSDFGMARL-MSADTHLSVSTLAGPGVYPE 1056
Db      889 HHDVPAIHHDDVKNLILDSKYEALIDFGLAKLMNSPNYHNMSRYA-----E 940
Qy      1057 YVOSFRCTKGDVSYGVVLELTGKRPDSDPFGDN-NLVGVYQAHAKLARSIVFDE 1115

```

---

```

Db      941 YGYTMNITEKSDVSYGVVLELTLSGRSAVE-PQIGDGLHVEWKKR-----MGTFEPA 994
Qy      1116 LMKEDPLF-----TELIQHLKVAACLDRAMRRPMVQVMAFKRQAGS---GID 1165
Db      995 LSVLDVLTQQLPDQIVQEMQLTGIMAFNCVNSPYERPTMEKVTVLLNEVKSPEEMKKT 1054
Qy      1166 SQGTIR 1171
Db      1055 SQPLIK 1060

RESULT 14
T10659
Probable serine/threonine-specific protein kinase (EC 2.7.1.-) T5F17.100 - Arabidopsis t
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T10659
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10659
A:Molecule type: DNA
A:Residues: 1-1013 <BEV>
A:Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.100
A:Experimental source: cultivar Columbia; BAC clone T5F17
C:Genes:
A:Gene: ATSP:T5F17.100
A:Map position: 4
A:Inserts: 876/1
C:Superfamily: protein kinase Xa21, leucine-rich alpha-2-glycoprotein repeat homology; p
C:Keywords: phosphotransferase; protein kinase

Query Match
Best Local Similarity 31.1%; Score 1313; DB 2; Length 1013;
Matches 377; Conservative 190; Mismatches 397; Indels 248; Gaps 39;

Qy      9 LSVTLTF-FFSEFSLSPQASPSQSLYREIHQILSPKQVLPDK-NLTPDW--SSNKPCTP 64
Db      3 MKIIVLFYCYIGSTSSVLASINDVNEBLSVSKTVLDPNLFKQMLSDTSDHCW 62
Qy      65 DQVTCRD-KYTSIDLSKPLNVFSAVSSSLTGLTSLFLSNHNGSVSGFKSCAS 123
Db      63 TGVNRNSGNVEKDLAAG-----NMLTG-----KIDPSIQL--SS 96
Qy      124 LTSLDLSNLSGPTTTLTSLGSCGLKFLVSSNTLDPFGKVSQGLKLNLEVLDSLAN 183
Db      97 LVSEFI-----SCNGFESLIPKS-----IPLKSIDISON 126
Qy      184 SIGANVGVWLSDCGELKHLAISGKISGDV--DYSRCVNLFLDVSNNFSTGIP-F 240
Db      127 SPSGSL--FLFSNESLGLVHLNLSGNNLSGNLTEDLGNVLSLEVDLDAGNPFQSLPS 183
Qy      241 LGDCSALQHLIDISGNLSDGFSRASTCTELKLNISNQFVPIPP--PKSLQYSL 298
Db      184 FKNIQKRFGLSGNNLTGLPVLQGLPSELTAIIGVYEFKGPPIPERGININSLKYL 243
Qy      299 AENKFTGEIPDPLSGACDTLTGLDLSGNHFGYAVPPFGSGSLSLSSNNFSGELPM 358
Db      244 AIGLSEIPSEL-----GKLSLETLILYENNFETGTPR 278
Qy      359 DTLAKGKLYLDSNERSGELPESLTVLSALTLTLDSSNNSGPIPLMLCONPQNTL 418
Db      279 E-IGSITTLVLFSDALTGELIPMEITKL-----KQ-L 310
Qy      419 QELVYONNGFTGKIPTLSNCSSELYSLHSFNYSGLTIPSSIGSLSKLDKMLNLE 478
Db      311 QLNLMNKKLSGSIIPRISLQVLQVLEIMNNTLSGEPSDLGKNSPLQWLDVSSNSPSG 370
Qy      479 EIPQELMYVTLETLIDFNLTGELIPSGLSNCTNLMNISNNRNGELIPKMTGR 538
Db      371 EIPSTLCKNKNLTKLILFNNTFGQIPATLSTCGSLVRYMGNLNSIPIPGKLEKL 430
Qy      539 ALLKSNNSPSGNIPELDDCSLITLMDLNTNLNFTGTPAAMFK-----QSGKIANFIAG 594

```

```

Db 431 QRIELAGNLSGIRPDIDSDVSLSPIDSRKQIRSSSTLTSLHINQAFIVADNFISG 490
Qy 595 KRYVYVYIKNDGMKKECHGAGNLLFQGIIRSEQJNLRSTNPNCTIRSVYGGHISPFEDNNG 654
Db 491 E-----VPDFQOCF 500
Qy 655 SMMFLDMSTYMLSGYIKRIGSMFYLFIINLGHNDISGIPDEVDLRGLNILDLSNKL 714
Db 501 SLNLDLSNNTLTGTPSSIASCEKVLNLRNNLTGIPQITMTSLAVLDLSNNSL 560
Qy 715 DGRIPQMSALMTLLEIDLSNNNLSGPIPEMGQFETPPAKFLNPGLCGYLPFCDPSN 774
Db 561 TGVLPESIGSPALLETLLNYSNKLTPVPINGPLKTNPDJLRGNSGLCGGYLPFCSKFO 620
Qy 775 ADGVYAHQSHGRPPAS-----LAGSVAMGLFSPFCIGLILVGBEMKRRRKEAELE 829
Db 621 R-ATSSHSHSGKRIYAGMLIGIASVLAIGI-----LTVVTLTKK----- 661
Qy 830 MYABGHNSGDRANTNTNMTLTVGKEALSINLAPEKPLKTLFADLLQATNGFHNDSLI 889
Db 662 WYSNGF--CGDELTASKGEMW-----RLMAFHR--LGFIASDILAC--IKESNMI 705
Qy 890 GSGGFGDVYKALIKDGSAY-AIKKL-----IHVSQGDREPAEMETTKIKRNLVPL 942
Db 706 GMGATGIVYNAEMGRSSTVLAVKLMRSADIEDGTTGD--FVGEVNLTKLGHNRNIVRL 763
Qy 943 LGYKRVGDEHLVNEVWKYSGLEVDLQDPKKG-VTKLSTRRIATGARGLAFLHNC 1001
Db 764 LGFLYNDKNNMIVYEFMLNGLGAIHGKRNAGGLLVDSWRNITALGVANGLAYLHHC 823
Qy 1002 SPFLIHDMKSNVLDLNTENLARYSDFGMARLMSAMDTHLSVSTLACTPGYVPEYQSF 1061
Db 824 HPEYTHDIDKNNLIDANLADARLADFGIARMMARKKE--TVSWAGSYGVIPEYGYTL 881
Qy 1062 RCRSTKGVYSGVVLLELTGKRPTDSDPDGDN-NLVGWKQAKKLIS--DVEDPEL-- 1116
Db 882 KYDEKIDIVSYGVVLELTGRRPLE-PEFGESVDIVAEVWRKTRNDISLEALDPVNG 940
Qy 1117 ---MKEDALLETLLQHKVAVACLDREARFPTVQVYAMFKEI--QAGSGIDSOSTIR 1171
Db 941 CRYVOE-----EMLVQIALCTTKLPKDRPSRDVYSMLGKPKRKSNSMEENTSR 994
Qy 1172 SIED---GGFST 1180
Db 995 SLAEKHSVST 1006
Qy 15
Db 15
receptor-like protein kinase 5 (EC 2.7.1.-) precursor - Arabidopsis thaliana
N.Alternate names: protein F2009.180
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jun-1999
C.Accession: S27756; T04620
R.Walker, J.C.
submitted to the EMBL Data Library, February 1992
A.Description: Receptor-like protein kinase genes of Arabidopsis thaliana.
A.Reference number: S27754
A.Accession: S27756
A.Molecule type: mRNA
A.Residues: 1-999 <MAL>
A.Cross-references: EMBL:M84660; NID:g166849; PID:AAA2859.1; PID:g166850
A.Experimental source: strain Columbia
R.Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohnle, J.; Mewes, H.W.; Meyer, K.F.X
submitted to the Protein Sequence Database, October 1998
A.Reference number: Z15380
A.Accession: T04620
A.Molecule type: DNA
A.Residues: 1-999 <BEV>
A.Cross-references: EMBL:AL021749
A.Experimental source: cultivar Columbia; BAC clone F2009
C.Genetics:

```

```

A.Gene: RLKS
A.Map position: 4
A.Intons: 868/1
A.Note: F2009.180
C.Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; P
C.Keywords: ATP; signal sequence #status predicted <SIG>
F.1-20/Domain: signal sequence #status predicted <SIG>
F.21-99/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>
F.66-89/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR2>
F.90-114/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
F.115-139/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>
F.140-163/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
F.164-187/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>
F.188-212/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>
F.213-236/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR8>
F.237-260/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR9>
F.261-284/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F.285-307/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F.308-331/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F.332-355/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F.356-379/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F.380-403/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F.404-427/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F.428-451/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR17>
F.452-475/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR18>
F.476-499/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR19>
F.500-523/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR20>
F.524-547/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR21>
F.548-570/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR22>
F.571-593/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR23>
F.681-971/Domain: protein kinase homology <KIN>
F.689-697/Region: protein kinase ATP-binding motif
F.98-102,150,269,282,452,576/Binding site: carbohydrate (Asn) (covalent) #status
F.711,737,819,821/Active site: Lys, Glu, Asp, Lys #status predicted
F.824,828/Binding site: magnesium (Asn, Asp) #status predicted

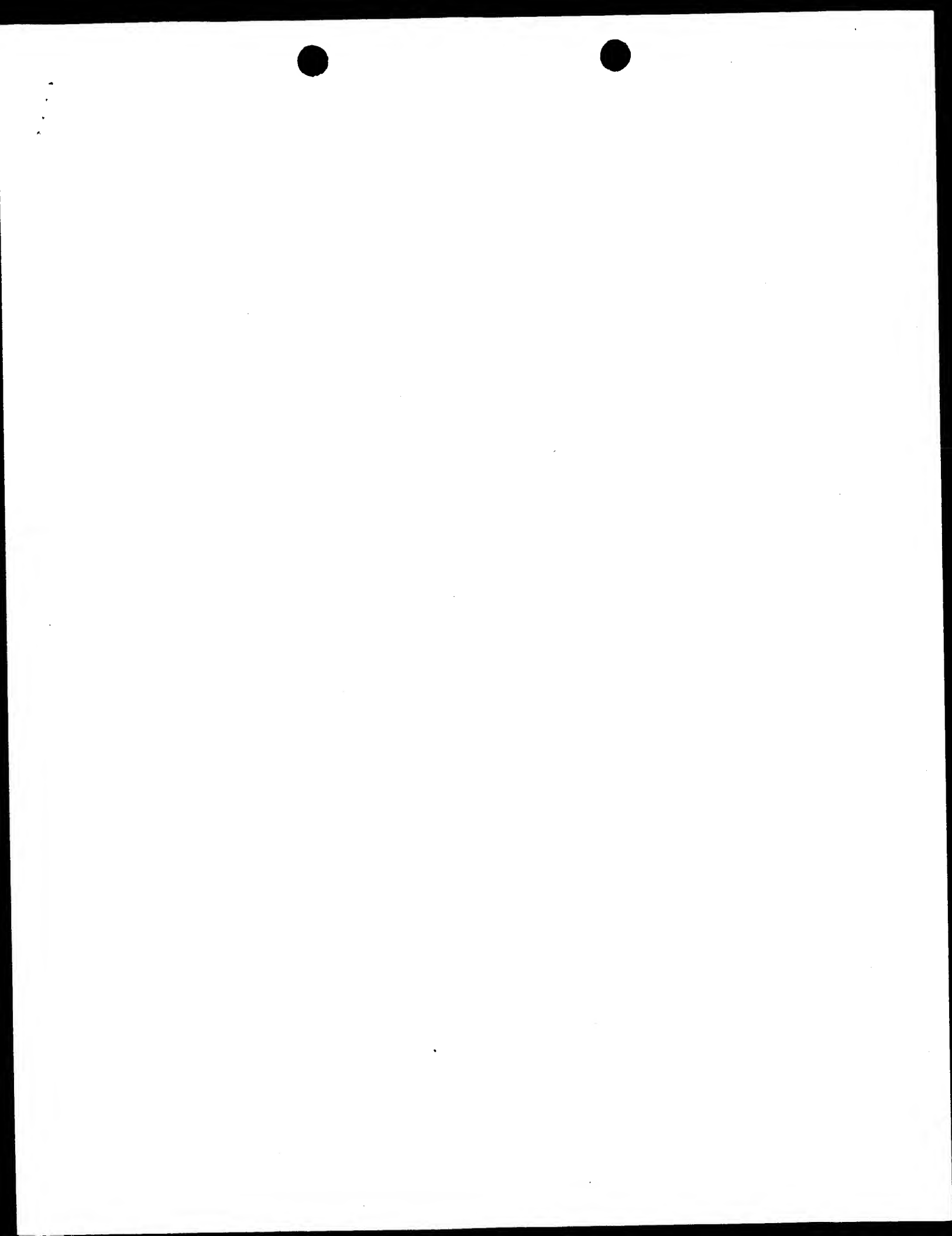
Query Match 21.2%; Score 1311; DB 1; Length 999;
Best Local Similarity 32.4%; Pred. No. 2.4e-52;
Matches 386; Conservativity 161; Mismatches 392; Indels 254; Gaps 30;

Qy 22 LSPQASPSQSLYREHQLISFQVLPD-KNLLPMSNKN--PCTFDGVC-RDDKVTSI 77
Db 11 LSTYLPSTLNQDATTIRQAKGLSDPAQSLSSMDNDVTPCWGLGVSCDANSNVSV 70
Qy 78 DLSKPLNVGFAVSSSLTSLTGLSFLSNHNSVSG--FKCSASLTSLDSRSL 135
Db 71 DLSKPLNVGFPSSI---LCHLPSTSLSYNNSINGSISADFDTCNLTSLDSENLV 127
Qy 136 GPVYTLTSLGSCSGKLFNLVSSNTLDFPKVSGGKALNSLEVLDSANSISGANVGVWL 195
Db 128 GSIP-----KSLPFLP----- 139
Qy 196 SDCCGSLKHLAISGNKISGDVDVSRVNLFLPVSNNFSTGIP-FLGDCSALQHLDISG 254
Db 140 -----NKLFLIEGNNISDTPSPSFGFRKLSLEINLAG 172
Qy 255 NKLSDFSRAISTCTELKLNISNOF-VGPIPLP--PLKSIQYSLAKNTGTGIPPL 311
Db 173 NPLSGTIPSLAGVTTLKEKELAYNLFSSQIPSOQAGNLTQVYMLACNLVGPPLPSL 232
Qy 312 SGACDTLTGLDLSGNHFGYAVPPFSSCSLLBSIALSSNFGSELPMDTLLKMRGLKVD 371
Db 233 S-RUTSLVNDLTFENQUTGSPWITQLKTVEOIEFLNNSFSELP-BMGWNTTLKRPD 290
Qy 372 LSPFESFGEPLPESLTMLASLTLTDLSSNFGPILPNCQNPKNLTQELVYQNGCFQK 431
Db 291 ASNNKLTGKLPDNLNL--NLBSLNFENMLBEGPLPESITRS--KTLSEIKLFNNRLTV 346
Qy 432 IPTLNSCSELVLSHFNYSLSGTIPSSLSGLSKADIKLMLMGEIPQELMYVYKLE 491
Db 347 LPSQAGNSPLQYVDLSYVNFSSGSEIPANVCGEKLLEYLLIIDSFGESGSLNIGKKSILT 406
Qy 492 TLIDPNDNGEIPSGLSNCTNMLWISLNNRLTGEIPKVIQGLLENLAIILKSNNSFSGN 551

```

Db 407 RYRLSNKLSGOIPHFPGWGLPRLSLLELSDNSFTGISIPKTIIGAKNLISNLRISKNRFSGS 466  
QY 552 IPDELGDCSRLLIWLIDNTNLFNGTI PAAMFKOSGKI ANFLAGKRYVYIRKDGKKECHG 611  
Db 467 IPNEIGSLNGIIEIGGAENDFSGEIPESI----- 495  
QY 612 AGNLEFGOIRSEOLNRLSTNPNCTISRYGHTSPPTDNGSMFELMSYNNLSGYIP 671  
Db 496 -----VKLKOLSR-----LDLSKNLSGEIP 516  
QY 672 KEIGSMPLYFIINTLGNHDSIPDEVDLARGNLIDLSNKLDRIPOMASALTMLTEI 731  
Db 517 RELRGKMLNELANNLNLSGEIPKEVGILPVLNVLIDLSNQSFGSIPLEIOLN-KLNLV 575  
QY 732 DISNNLSGPIPENGQFTEPPAKFLNPNGLCYPLPCDPSNADGYAHHQSHGRPPAS 791  
Db 576 NLSYNHLSGKIPLPYANKIY-AHDFIGNPGLC-----VDLDGLCR-----K 615  
QY 792 LAGSVAMG-----LLFSFVCLFGLILVGRMRKRRKKEALEMYAEGHNSGDRT 842  
Db 616 ITRSKNITGYWILITFLAGLVFVVGIVMEIACRKLRLAKSSTLAA----- 663  
QY 843 ANNTNMKLTGYKALISINLAFAFERPLRKLTF-----ADLLQATNGFHNDLSIGSGGFQDV 897  
Db 664 ---SKM-----RSFHLKHFSEHEIADCLDERN-----VIGFGSSGRV 697  
QY 898 YKALLKXGSAVAIKKLIHVSQGDRE-----FMAEMETIGIKIRNLVPLIGYCK 947  
Db 698 YKVELRGGEVVAVKKLNKSVYKGDDEYSSDLNRDVAFAEVEITLGTIRHKSIVRLWCCCS 757  
QY 948 VGERLILVNEVMKYGSLEVDYLOPKKGGVLYKLSRKTALISARGLAFLHNCSPHIY 1007  
Db 758 SGDCKLIVYEMRPGSLADVLHGDRKGGVLYGMPERLIALDNAEGLSYLHDCVPPIVH 817  
QY 1008 RDMKSSVLLDENLEAVSDFGMARL--MSAMDTLHLSVSTLAGTPGYVPEYYOSFRCT 1065  
Db 818 RDYKSSNILLDSYGAADFGIAKVGOMSGSKTPREAMSGIAGSCYIAPERYVTLRYNE 877  
QY 1066 KGDVYSYGVVLELLTGKRPDSDPDGNNLVGVV-KQHAKLISDVPDELMKEDPALE 1124  
Db 878 KSDIYSFGVLLLELVGKOPTDS-ELGDKMAKWCCTALDKCGLFVIDPKL--DLKFK 933  
QY 1125 IELLQHLKVAVACLDRAMRPTMVOVMAMFKEIQAGSGIDSGTIRSIEDGG 1177  
Db 934 EELSKYVHIGLCTSPPLNRPBMRKVIMLQEVGAVPCSSPNTSKSKXTGG 986

h completed: March 10, 2003, 18:24:52  
time : 41.2258 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 17:58:34 ; Search time 25.6693 Seconds

(without alignments)  
1964.817 Million cell updates/sec

Title: US-09-823-394-2  
Sequence: 1 MKTFFSFFLSVTLTFFSF...GRSTIEMVMSIKVEPDEGL 1196

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 188354 seqs, 42170167 residues

number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

1: /cgnt2\_6/ptcodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgnt2\_6/ptcodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgnt2\_6/ptcodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgnt2\_6/ptcodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
5: /cgnt2\_6/ptcodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
6: /cgnt2\_6/ptcodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgnt2\_6/ptcodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
8: /cgnt2\_6/ptcodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgnt2\_6/ptcodata/1/pubpaa/US09\_PUB.pep.\*  
10: /cgnt2\_6/ptcodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
11: /cgnt2\_6/ptcodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
12: /cgnt2\_6/ptcodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgnt2\_6/ptcodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgnt2\_6/ptcodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6183	100.0	1196	US-09-823-394-2	Sequence 2, Appl1
2	1284.5	20.8	960	US-08-910-386A-18	Sequence 18, Appl1
3	1087	17.6	1025	US-08-910-386A-7	Sequence 7, Appl1
4	1055	17.1	1161	US-08-910-386A-20	Sequence 20, Appl1
5	1009.5	16.3	996	US-08-910-386A-5	Sequence 5, Appl1
6	937	15.2	854	US-09-754-853A-1111	Sequence 1111, Ap
7	937	15.2	854	US-09-754-853A-1113	Sequence 1113, Ap
8	937	15.2	854	US-09-754-853A-1115	Sequence 1115, Ap
9	937	15.2	854	US-09-754-853A-1117	Sequence 1117, Ap
10	937	15.2	854	US-09-754-853A-1119	Sequence 1119, Ap
11	937	15.2	854	US-09-754-853A-1121	Sequence 1121, Ap
12	933	15.1	854	US-09-754-853A-1123	Sequence 1123, Ap
13	933	15.1	854	US-09-754-853A-1125	Sequence 1125, Ap
14	933	15.1	854	US-09-754-853A-1127	Sequence 1127, Ap
15	933	15.1	854	US-09-754-853A-1129	Sequence 1129, Ap
16	933	15.1	854	US-09-754-853A-1131	Sequence 1131, Ap
17	933	15.1	854	US-09-754-853A-1133	Sequence 1133, Ap
18	933	15.1	854	US-09-754-853A-1135	Sequence 1135, Ap
19	933	15.1	854	US-09-754-853A-1137	Sequence 1137, Ap

20	933	15.1	877	9	US-09-754-853A-1104	Sequence 1104, Ap
21	933	15.1	877	9	US-09-754-853A-1106	Sequence 1106, Ap
22	932	15.1	854	9	US-09-754-853A-1108	Sequence 1108, Ap
23	932	15.1	877	9	US-09-754-853A-1109	Sequence 1109, Ap
24	822.5	13.3	894	9	US-09-754-853A-1099	Sequence 1099, Ap
25	822.5	13.3	894	9	US-09-754-853A-1105	Sequence 1105, Ap
26	822.5	13.3	894	9	US-09-754-853A-1111	Sequence 1111, Ap
27	821.5	13.3	894	9	US-09-754-853A-1117	Sequence 1117, Ap
28	817.5	13.2	894	9	US-09-754-853A-1119	Sequence 1119, Ap
29	763	12.3	813	8	US-08-910-386A-9	Sequence 9, Appl1
30	656.5	10.6	612	8	US-08-910-386A-2	Sequence 2, Appl1
31	629.5	10.2	749	10	US-09-772-134B-14	Sequence 14, Appl1
32	617.5	10.0	731	9	US-09-828-113-27	Sequence 27, Appl1
33	614.5	9.9	731	9	US-10-086-464-17	Sequence 17, Appl1
34	604.5	9.8	647	9	US-10-086-464-2	Sequence 2, Appl1
35	604.5	9.8	647	9	US-10-086-464-4	Sequence 4, Appl1
36	604.5	9.8	721	9	US-10-086-464-5	Sequence 5, Appl1
37	601	9.7	633	9	US-10-086-464-11	Sequence 11, Appl1
38	595.5	9.6	674	9	US-10-086-464-14	Sequence 14, Appl1
39	586	9.5	731	9	US-10-086-464-8	Sequence 8, Appl1
40	581.5	9.4	355	10	US-09-862-027-26	Sequence 26, Appl1
41	505	8.2	604	10	US-09-862-027-17	Sequence 17, Appl1
42	499	8.1	645	10	US-09-862-027-15	Sequence 15, Appl1
43	499	8.1	645	10	US-09-862-027-16	Sequence 16, Appl1
44	482.5	7.8	339	10	US-09-838-955-8	Sequence 8, Appl1
45	481	7.8	669	10	US-09-862-027-27	Sequence 27, Appl1

## ALIGNMENTS

RESULT 1	US-09-823-394-2	Application US/09823394
1	Sequence 2, Appl1	US-09-823-394-2
2	Publication No. US20030041344A1	
GENERAL INFORMATION:		
APPLICANT:	Chory, Joanne	
APPLICANT:	Jianming, Li	
TITLE OF INVENTION:	Salik Institute for Biological Studies	
FILE REFERENCE:	SALKINS 012CP1	
CURRENT APPLICATION NUMBER:	US/09/823,394	
CURRENT FILING DATE:	2001-03-30	
PRIOR APPLICATION NUMBER:	08/881,706	
PRIOR FILING DATE:	1997-06-24	
NUMBER OF SEQ ID NOS:	2	
SOFTWARE:	FastSeq for Windows Version 4.0	
SEQ ID NO 2		
LENGTH:	1196	
TYPE:	PRT	
ORGANISM:	Arabidopsis	
US-09-823-394-2		
Query Match	100.0%;	Score 6183; DB 9; Length 1196;
Best Local Similarity	100.0%;	Pred. No. 0;
Matches 1196; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MKTFFSFFLSVTLTFFSFSLSPQSPQSLYRETHQISPDVY.PDKRLPDWSSNKK 60
DB	1	MKTFFSFFLSVTLTFFSFSLSPQSPQSLYRETHQISPDVY.PDKRLPDWSSNKK 60
QY	61	PCTPDGVTCRRDKTYSIDLSKPLNVGFSVSSLSLTGLBSLFLSNHINGSVGFKC 120
DB	61	PCTPDGVTCRRDKTYSIDLSKPLNVGFSVSSLSLTGLBSLFLSNHINGSVGFKC 120
QY	121	SASLTSLDLSRNSLSGVTTLTSLGSCGLKFLNVSSNTLPDPGRKVSGLKXNSLEVDL 180
DB	121	SASLTSLDLSRNSLSGVTTLTSLGSCGLKFLNVSSNTLPDPGRKVSGLKXNSLEVDL 180
QY	181	SANISGANVGVWLSDGCELKHALISGNKISGDVDVSRVNLFLDVSSNNFSTGTFP 240
DB	181	SANISGANVGVWLSDGCELKHALISGNKISGDVDVSRVNLFLDVSSNNFSTGTFP 240

```

QY 241 LGGCSALQHLDISGNKLSGDFSRATSTCTELKLNLSNQFVGPPIPPULKSLQYLSLAE 300
D 241 LGGCSALQHLDISGNKLSGDFSRATSTCTELKLNLSNQFVGPPIPPULKSLQYLSLAE 300
QY 301 NKFTGPIPPFLSGACDTLTGLDISGNHFGAVPPFPFGCSLLBSLALSSNNSSGGLPMDT 360
D 301 NKFTGPIPPFLSGACDTLTGLDISGNHFGAVPPFPFGCSLLBSLALSSNNSSGGLPMDT 360
QY 361 LLMRGKLVLDLSFNEFSGELPSLNTLSASLLTLDLSNNFSGPILPMLCONPKXTLOE 420
D 361 LLMRGKLVLDLSFNEFSGELPSLNTLSASLLTLDLSNNFSGPILPMLCONPKXTLOE 420
QY 421 LYLQNGFTGKIPTLNSCELSVLSLSPNYLSGTTSSLSGSLSKRLDIKLMNLEGGI 480
D 421 LYLQNGFTGKIPTLNSCELSVLSLSPNYLSGTTSSLSGSLSKRLDIKLMNLEGGI 480
QY 481 POELMYVKTLETLILDNDLTGELPSGLSNCNTLMNLSNNRLTGEIPKWLLENLAI 540
D 481 POELMYVKTLETLILDNDLTGELPSGLSNCNTLMNLSNNRLTGEIPKWLLENLAI 540
QY 541 LKLSNNSFSGNIPDELDCRSLIWLDTNTLFGTTIPAAFKOSGKIAANFIAGKRYVYI 600
D 541 LKLSNNSFSGNIPDELDCRSLIWLDTNTLFGTTIPAAFKOSGKIAANFIAGKRYVYI 600
QY 601 KNDGKKECHGAGNLLEFGIRSFQNLSTRNPCNITSRYVGGHTSPPTDNNKSMWFLD 660
D 601 KNDGKKECHGAGNLLEFGIRSFQNLSTRNPCNITSRYVGGHTSPPTDNNKSMWFLD 660
QY 661 MSYNNLSGYIPEKIGSMPLYFLNLGNHNDISGSIPEDEVGLRGINILDLSSNKLDRIPQ 720
D 661 MSYNNLSGYIPEKIGSMPLYFLNLGNHNDISGSIPEDEVGLRGINILDLSSNKLDRIPQ 720
QY 721 AMSALTMLTEIDLSNNNLSGPIPEWGFETPPPAKFLNPGLCYPLPCDPSNADGIAH 780
D 721 AMSALTMLTEIDLSNNNLSGPIPEWGFETPPPAKFLNPGLCYPLPCDPSNADGIAH 780
QY 781 HORSGRBPASLAGSVAMGLLPSFYCTFGLILVGRBEMKRRKKEALEMYAEGHNSGD 840
D 781 HORSGRBPASLAGSVAMGLLPSFYCTFGLILVGRBEMKRRKKEALEMYAEGHNSGD 840
QY 841 RTANNTNKLTCVKALSLINLAFAEKPLRKLTFADLLQATNGFHNDSLISGGFGDYKA 900
D 841 RTANNTNKLTCVKALSLINLAFAEKPLRKLTFADLLQATNGFHNDSLISGGFGDYKA 900
QY 901 ILKDSAAVAIKKLHVSGQGRREFMAEMETIGKIKRMLVPLAGYCKVGDRLLENVWK 960
D 901 ILKDSAAVAIKKLHVSGQGRREFMAEMETIGKIKRMLVPLAGYCKVGDRLLENVWK 960
QY 961 YGSLDEVLODPKKGKVKLSTRRKIAIGSANGLAFLHNNCSPHIHRDMKSSNVLDEN 1020
D 961 YGSLDEVLODPKKGKVKLSTRRKIAIGSANGLAFLHNNCSPHIHRDMKSSNVLDEN 1020
QY 1021 LBARVSDFCMARLMSAMDTHLSVSTLACTPGYVPEYYSQFPCSTKGDVYSYGVLLLEL 1080
D 1021 LBARVSDFCMARLMSAMDTHLSVSTLACTPGYVPEYYSQFPCSTKGDVYSYGVLLLEL 1080
QY 1081 TGRKPLTSDPDGDNNUYGVWVQKAKRLISDVDPBELMEDPALLELOHLKVAACDD 1140
D 1081 TGRKPLTSDPDGDNNUYGVWVQKAKRLISDVDPBELMEDPALLELOHLKVAACDD 1140
QY 1141 RABRPLTVMQVMAFKEIOAGSGIDSGSTIRSIDGSGFTIEMVMSKEVEGKL 1196
D 1141 RABRPLTVMQVMAFKEIOAGSGIDSGSTIRSIDGSGFTIEMVMSKEVEGKL 1196

```

RESULT 2  
US-08-910-386A-18

; Sequence 18, Application US/08910386A  
; Patent No. US20020092041A1  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Pamela C.  
; APPLICANT: Wang, Guo-Liang  
; APPLICANT: Song, Wen-Yuang

```

APPLICANT: Hulbert, Scot
APPLICANT: Richter, Todd
TITLE OF INVENTION: Procedures and Materials for Confering
TITLE OF INVENTION: Disease Resistance in Plants
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,386A
FILING DATE: 13-Aug-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baebian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-058950US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-386A-18

Query Match 20.8%; Score 1284.5; DB 8; Length 960;
Best Local Similarity 32.1%; Pred. No. 9.3e-75; Indels 257; Gaps 35;
Matches 378; Conservative 168; Mismatches 373;

QY 10 SVTTLFFPFFSLFQASPSQ---SLYREIHQILSFQDVLPDKNL---LPDW--SSNKN 60
D 4 SNTSLFFAYFLVFLITPSQSRNLSLRQAKTVLSKYAFVQSSVPSTLSNMNNSNYS 63
QY 61 PCTFDGVTGRDK-VTIDLSKPLNVGSAVSSLSLTGLLESPLNSHNGSVSGFK 119
D 64 ICSWGITCDDTKSVTSIDIS-----NLNISG-- 90
QY 120 CSASLTSLDSRNSLSGPTTTLTSLGSCGKLFELNVSSNTLDPGKVGGLK-LNSLEVL 178
D 91 -----SLSFDIHELTR-----LRVAINSNL--FSGNLGWEYREFFVLQVL 129
QY 179 DLANSISGANVYGVWVLSGCGELKHLAISGNKISGDVDVSRCVNLEFLDVSSNNFSTGI 238
D 130 DAYNNPFSGLPLPG-----VTQLVQDKYINFGNYSFGKI 164
QY 239 PF-LGGCSALQHLDISGNKLSGDFSRATSTCTELKLNLS-SNOFGAPLPV--PLKSPQ 294
D 165 PLGYGFMQLEFLSLAGNDLHGPIPRELGNVTSLRWLQGYNQFDEGIPPELGKLVNLY 224
QY 295 YLGLAENKFTGELIPDFLSGACDTLTGLDISGNHFGAVPPFPFGCSLLBSLALSSNNFSG 354
D 225 HLDSSCNLTGSLPPPL-GNLNMLDTLFLQXQLTGVFPFQGLNLRKLSLDISVETLG 283
QY 355 ELPMDTLLKMRGLKVLDSLNEFSGELPSLNTLSASLLTLDLSNNFSGPILPMLCONPK 414
D 284 EIPVD-LSGKELILNLFLINNHLGELPGCIAEL-PLEMLNLMRNKFTGSLPSKLGMMG 341
QY 415 KNTLQELYLQNGFTGKIPTLNSCELSVLSLSPNYLSGTTSSLSGSLSKRLDIKLMN 474
D 342 K-LIEIDLSNRLTGLIPSLCGRNKLTLILDNLFLEPLDDPGQCRTRLSRVAMGN 399
QY 475 MLEGEIPQELMYVKTLETLILDNDLTGELIPSGLSNCNTN-LMNLSNNRLTGEIPKWL 533

```

```

Db 400 YLSSGIPFGPIYELBELSVLELONNYISGOLWNEKSSASSKLEGLNLSNNRLSGALPSAIG 459
Qy 534 RLENIATIKLSNNSPFGNIDELDCRSLIWLIDNTLNFNGTIPAMAFKQSGKIANPFA 593
Db 460 NYSGLKMLVLTNGNFGSDIPEDIGRLSKLIDLSRNPFSGTIP----- 503
Qy 594 GKRVYIKNDMKKECHGAGNLEFQGRSEQLNLSRNPNCNITSRYVGHGHTSPFDNN 653
Db 504 -----PQIGNC 509
Qy 654 GSNMFLDMSYNNLSGYIPKEIGSNPYFLINLGHNDISGIDPEVDLACINILDLSSNK 713
Db 510 LSLTYLDSQNGSLGPIVQIQAQHILNINISWNHFNESLPELIGLMS----- 559
Qy 714 LDGIPQMSALTMLEITIDLSNNLSGPIPEMGOFEFPAPKLNPGLCGYLPBCDPS 773
Db 560 -----LTSADFSHNNLSGSIPEQTQYLNFNSTPTNGPYLSG---SDSTPS 602
Qy 774 NADGYAHORSHG-----RRPASLGSVAMGLLFSFVCTFGILVGRERKRKRKEAEL 828
Db 603 NITSNPSSELDGSDSRKTKVPTTKTFAPGLTF---C-SLIFVLAITKTRKSGK--- 654
Qy 829 EMTAGHSGNSDRTANNNTWKLTGVKEALSIMLAPEKELRKLTPA--DLQATNGFHND 886
Db 655 -----NSNMLKLT-----AFQ-----KLEFGSDVLCQ---LIDN 681
Qy 887 SLIGSGFGDVYKAILKQSAVAIKKLHVSQGGDEFEAEMETIGKIKNRYVLLGYC 946
Db 682 NVLGRGAGIVYGTMPNDHVAVKLGISKSHDGLSAELNTLGKIRHRYVALLAFC 741
Qy 947 KYDDEBLVNEVWKYSLBDVLQDPKGGVVKLTSTRRTIAGSARGLAFIHNHNSPHIT 1006
Db 742 SNKEVNLVYETWNLNSLEVLHG--KNGGQLOMETRLKIALEAKGLSYLHHCSPWII 799
Qy 1007 HRMKSNNVLDENLEFARYSDFGMALMSAMDTHLSVTLAGTPGYPEYOSFPGSTK 1066
Db 800 HRDVKSNNTILNSELFAHVADEFLAKYFNNGTSECMASLNGSYIAPETAYLTKDEK 859
Qy 1067 GDVYSYGVVLELTLGKRPTDPPDGDN--NLVGVNKHAKLR--ISDVPPELMKEDP 1121
Db 860 SDVYSYGVVLELTLGKRPTDPPDGDN--NLVGVNKHAKLR--ISDVPPELMKEDP 1121
Qy 1122 ALEIELQHLKVAVALCLDDRAMRREPTMYQVAMAFKE 1157
Db 916 AI-VEAMQVFVAMLCVEEYSIERPTMREBVOMLSQ 950

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,386A
FILING DATE: 13-AUG-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baebian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-058950US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1025 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-386A-7

Query Match 17.6%; Score 1087; DB 8; Length 1025;
Best Local Similarity 30.8%; Pred. No. 5e-62;
Matches 323; Conservative 140; Mismatches 371; Indels 216; Gaps 28;

Qy 231 SNFSTGI--PFLGDSALQHDHISGNKLSGDFSRATSTELKLNINSQVFGPIPL 288
Db 88 SSNLS-GILSPGLNLSFLRLDLDNVLSEIPELSRLSLQLESDNSIGSIPA- 145
Qy 289 PLKSLQYSLANKETGEIPDFLSGACDTLTGLDLSGNHFGAVPPEFG-SCSLSEIAL 347
Db 146 -----ATGACTKLTSLDLSHNLQNGMIPREIGASLKLKLSLYL 183
Qy 348 SSNFSGELPMDTLKMGKLVLDLSPNEPSELPESTLNSASLTLDLSSNNSGPIL 407
Db 184 YKNGLSGEIP-SALGNLTSLOEPDLSTFRSLGALPSLGLS-SLTMNNGQNNLSG-MI 240
Qy 408 PVLCON-----PKVTLQELVY-----QNNGFTKIPPTNSGSELY 443
Db 241 PMSINWLSLRAFSVRENKLGMIPTNAFTLHLEVIDMGTRFHCKTIAVSANSHLT 300
Qy 444 SLHSPNLSGTPISLSGLSKLDKLMNM----- 475
Db 301 VIQIGNLPFGITISGFRRLNLTLEYLMRNLPTREDDWGFISDLTNCISKLOTNLGE 360
Qy 476 --LSEELPOELMYKT--LETLIDFNDLGEIPSGNCTNLWISLNNRLTGEIPKI 532
Db 361 NNLGGLVLPNSPNSLSTSLFLALELAKITSGIPDIGNLIGLOHLYLCNNRFGSLPSL 420
Qy 533 GRLNLAILKLSNNSPFGNIPDELDCRSLIWLIDNTLNFNGTIPAMFKQSG---KIA 588
Db 421 GRKNTLGLLAYENNLGSIPLAIGNLTENILLLGNNKSGMIPTLSNLTNLSLGLS 480
Qy 589 ANFIAGKRYVYIKNDMKKECHGAGNLEFQGRSEQLNLSRNPNCNITSRYVGHGHTSP 648
Db 481 TNNLSG-----IPSELFNIOQLSIMINVSNNLSKNNLSGSIPO 515
Qy 649 TFDNNGSMFLDMSYNNLSGYIPKEIGSNPYFLINLGHNDISGIDPEVDLACINIL 708
Db 516 EIGHKMLVEFHAESNRLSGKIPNTIGDQQLRYLYLQNNLSGSIPLSLGQLKGLTLD 575
Qy 709 LSSNKLGRIPQMSALTMLEITIDLSNNLSGPIPEMGOFEFPAPKLNPGLCG--- 764
Db 576 LSSNLSGQIPSLADITMHLSTLSPNSFVGEVPTIGAFASAGISIQNAKLCGGIPD 635
Qy 765 YLPRLCDPSNADGYAHORSHGRPRASLAGSVAMGLLFSFVCTFGILVGRMRKRK 824
Db 636 LHLPRCP-----LLENRKGFPVLPISVSLAALAILSLYLI-----TWHKRTK 682
Qy 825 EAELEMTAGHSGNSDRTANNNTWKLTGVKEALSIMLAPEKELRKLTFADLQATNGFH 884
Db 683 GABSRITSMKH-----PL--VSQQLVATDQFA 709
Qy 885 NDSLIGSGFGDVYKAILKQSAVAIKKLHVSQGGDEFEAEMETIGKIKNRYVLLG 944

```



QY 1035 SAMDTHLSVSTAGTGYVPEYVYOSFRCTGADYVSGVLLLELTGKRPTDSDPDND 1094  
 Db 1048 GTSETH-ATTGVAAGTGYVAPYAMTCRAVDYSGVLLLELTSPKALD-PSFSSY 1105  
 QY 1095 ----NLVGV---VKOAKLRIS 1109  
 Db 1106 GNGFNIVAMACMLLRKAVLRSS 1127

## RESULT 5

US-08-910-386A-5  
 ; Sequence 5, Application US/08910386A  
 ; Patent No. US20020092041A1

GENERAL INFORMATION:  
 APPLICANT: Ronald, Pamela C.  
 APPLICANT: Wang, Guo-Liang  
 APPLICANT: Song, Men-Yuang  
 APPLICANT: Hulbert, Scott  
 APPLICANT: Richter, Todd  
 TITLE OF INVENTION: Procedures and Materials for Confering  
 TITLE OF INVENTION: Disease Resistance in Plants  
 NUMBER OF SEQUENCES: 53  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/910,386A  
 FILING DATE: 13-AUG-1997  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Baestian, Kevin L.  
 REGISTRATION NUMBER: 34,774  
 REFERENCE/DOCKET NUMBER: 023070-058950US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 996 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-910-386A-5

Query Match 16.3%; Score 1009.5; DB 8; Length 996;  
 Best Local Similarity 30.5%; Pred. No. 4,66-57;  
 Matches 310; Conservative 148; Mismatches 391; Indels 167; Gaps 29;  
 QY 161 DPGKVGSGKLINSLEVDLSANSI--SGANVVGWVLSDG-----CGELKH----- 204  
 Db 24 DDDGDAAG---DELALISFKSSLLYGGGSLASWNISGHCHQCTWGVVGR-RHHRV 78  
 QY 205 --LAISGKIGDGVAVSCVNLRF--LDVSSNNFSTGIP-FLGDCSALOHLIDISGKLS 258  
 Db 79 VKLRKRSNLTGIIISPS-LGNLSFKRTQLSNHLSGKIPQELSRSLQQLVNFNLSLS 137  
 QY 259 GDFSAISTCTELKLNINSNQFGRIPPL--PLKSLQYLSIAKNTKPTGEPDLSGACD 316  
 Db 138 GEIPALGNLTSLVLEIHNNTLSGTSISLQKLTGLYNLAENMLSGSIPTSF-QQLR 196  
 QY 317 TLTGIDSGNHYGAVPPFSGSLLESIALSNNFSGELPMDTLLKMRGLKAVDLSFNE 376  
 Db 197 RLSTLSLAFNHLGSAIPPIWNISLTIFFEVVSNNTLTLPANFNLPIQLQVPMYNN 256

QY 377 FSGELPESLTNLSA-SILTLDLSSNPNFSGIIPNLCONPKNTLOELEYL-----ONNG 427  
 Db 257 FHGPISASINASISIFITGL--NSFSGVVPPEI--GMRNLQRLLEPTELEAEFTND 312  
 QY 428 FTGKIPPTLSNCELYSLHSFNYLSGTIPSSIGSL-SYLRDIKMLNMLEBIPQELMY 486  
 Db 313 W--KFMALTNCNLTQEVLEIAGCKFGGVLPDSVSNLSSLSVSIKDNKISGSLPDIQN 370  
 QY 487 VKTLETLIDFNDLTGEIPSGLSNCTNLMISLNNRLTGEIPKMTGRENLAILKLSN 546  
 Db 371 LVNLQYISLANNLSFOSLPSFSFKLKNLRLTYDNNRLTGLPLTIGNLTQLTNNEVQFN 430  
 QY 547 SFGNTPDELGDCKSLIWLDTNLNFGTIPAMF-----KOSGLAANFIAGKRYVIK 601  
 Db 431 AFGTIPSTLGNLTGLFQNLGNHNFITGIPLEIFIPALSELIDVSHNNLG-----483  
 QY 602 NDMKKECHGAGMLRFQGRISQRLSTRNDCNITSRYVGHTSPFPDNNSGMFLDM 661  
 Db 484 --SIPKEIGKKNIVEFHA-----500  
 QY 662 SYNNLSGYIPKEIGSMPLYFLINLGHNDISGIPDEVDLRLNLTLSNKLDRIPQA 721  
 Db 501 DSKLGEIPSTIGECQLQHLFLQNNFLNGSIPALTLQKGLDITLDSGNNLSGQIPWS 560  
 QY 722 MSALTMLTBIDLSNNLSGPIPEMGQFETFPAPKFLNPNLGC---YPLPCDPSNADG 777  
 Db 561 LGDMTLHSLNLSFNSFHBGVPTNGVFANASEIYIQGNHICGIPELHLPTCS-----614  
 QY 778 YAHQSHGRPPASLAGSYAMGLFSFCIFGLILVGRMRKRRKKEALEMYABHGN 837  
 Db 615 ---LKSRRKRKHQIILLVWVICLVSTLAVFSLVWLTCHK-RRKEVPATTSMQG--667  
 QY 838 SGDRYANNNTWKLTVKALSLINLAPEKPLKLPADLQATNGHNDLSLGSFGDV 897  
 Db 668 -----PM--ITYKQVKAIDGSSSHLGSFSGV 696  
 QY 898 YKAIL--KDG---SAVAIKKLIVSGGDRFMAEMETIGIKRHNLYVLLGYCYKGR 952  
 Db 697 YKGFPSQGEITSLVAVKVLEKLPKALKSTFABEETLRNRHNVIVITICSIDNR 756  
 QY 953 -----LVNENWKKYSGLEVLDOPKKGVK--LKSTRRKIAISAGLAFLHNCSPH 1004  
 Db 757 GNDPKAIVYDFENGLLEBMLPBTNDQAEQRHLTHQVTLILVACALBHLHFHGE 816  
 QY 1005 IIRDMKSSNVLLDETLBARVSDPGMARLM---SAMDTHLSVSTLAGPGVPEYVYOS 1060  
 Db 817 IVYCDIKSSNVLLDPMVAHVDFGLARILVGGSSLMQOSTSSMGIRGITGAABEYGV 876  
 QY 1061 FRCTGADYVSYGVVLLLELTGKRPTDSDPDNDNLVGVVKOAKLRISDVPDEL 1116  
 Db 877 NTASTHGDYSGIVLETVTGMRPADSTFRGTGLSRQYVEGHLGRIMDVDRKL 932

## RESULT 6

US-09-754-853A-1111  
 ; Sequence 1111, Application US/09754853A  
 ; Publication No. US20030005491A1

GENERAL INFORMATION:  
 APPLICANT: Hauge, Brian M.  
 APPLICANT: Parnell, Laurence D.  
 APPLICANT: Parsons, Jeremy D.  
 APPLICANT: Wang, Ming Li  
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Soybean Cyst Nematode Resistance  
 FILE REFERENCE: 38-10(15810)B  
 CURRENT APPLICATION NUMBER: US/09/754, 853A  
 PRIOR FILING DATE: 2001-01-05  
 PRIOR APPLICATION NUMBER: US 60/174,880  
 PRIOR FILING DATE: 2000-01-07  
 NUMBER OF SEQ ID NOS: 1119  
 SEQ ID NO 1111  
 LENGTH: 854

```

; TYPE: PRT
;
; ORGANISM: Glycine max
;
; FEATURE:
;
; OTHER INFORMATION: Clone ID: rhg1_noir_amplicon
;
US-09-754-853A-1111

```

Query Match 15.2%; Score 937; DB 9; Length 854;  
Best Local Similarity 31.3%; Pred. No. 1.7e-52;  
Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24,

Qy	290	IKSLQVYLSAENKRETEBIPPLISGACDTLTGDLSDGNHYAVAPPPFSCSLIEBLS	349
Db	139	LQGRKSLSDHNOGGISPELT-GLPNLRGYVLPNNRLTGSIPSLSGFCLDGLSDLSN	197
Qy	350	NNFSGELPMDTLKMRGLKYLDSFNEFSGELPESLTNLISASLLTLJDSNNNSPILPN	409
Db	198	NLTGALIPY-SLANSTKLYMLNISFNSFGPLPALYTH-SFSLTFSLSDNNNLGGSLPNS	255
Qy	410	LCQPNK---TLQELTYQNNGFPGTKIPPLTNSCELSVLSLSPNLTSGTIPBSGSLK	466
Qy	256	WGSNSKGFPEPLQWLLIDNNPFPGDVPASIGSLREINETSLSHNKPSGAIPIEIGTSLRL	315
Db	316	K-----TLJDSNNALNENLPAITSLNISSLTLMAENNLLDN	351
Qy	527	BIPKMTGLENLAIKLKSNNSFSGNIPDELGDCRSLLIMDLTNLTNGTIPAMFGOSK	586
Db	352	QIPQSLGRLNLTSLILSRNOFGSHIPSSIANISSIRQLDSINNPSGEIPY-----	403
Qy	567	IAANPIAGKRYVYIKNDGMKKECHGAGNLEPQGISSEQLNRSLTRNPNCTISRYGHT	646
Db	404	-----	403
Qy	647	SPTFDNNGSMFELMSYNMLSGYIKPEIGSMPLYFILINGHNDISGJPDVEYDGLRINI	706
Db	404	--SFSQSRSLNLFVNVSTNSLSGSVP-----PLL-----	429
Qy	707	LDLSNNKLDGRIPOAMGALTMLTLEIDLSNNNLSGPIPMGQEPTEPPAKFLNPNLQCY-	765
Db	430	-----AKFFNSSFFVGNIOQCYX	448
Qy	766	PLPRC-----DPSNADGYAHNQRSHGRKRPASLAGSVAMGLLFFVYIIFGLIIVG	814
Db	449	PSTPLSQAPSGQVIAPPEVYSKHHHRKLTSDIILIVAGVLVLLILCCVLLFCIL-	507
Qy	815	REMRKRRKKAELMYAEGHNGSD-----RTAANTYMKLTGYEA---LSINIAAFE	865
Qy	508	---RKRSTK-----AGNGQATEGRAAATMTTEKGVPPVAAGDYBAGEGAKULYHFD	556
Db	866	KPLRLTAPADLLQATNGFHNDLSLIGSGFQDYYKAILKDGSAVALIKLILHVSQGDREPM	925
Db	557	GPM-AFTADDLICAT-----AEIWGSTYGVYKAILDEDSQYAVRLEKRTIKGHRFE	610
Qy	926	AEMETIGIKIRNVLPLLYGKVG--DBRLINNEVMKYGSLLEVYLDPPKKGVK--LKLS	981
Db	611	SEVSLYIGIRHPNVALALRAY-YLQPKGEKLLVPDMYSGSLASFLHG--GGTETFTOMP	666
Qy	982	TRRKIAISAGLAFVLHNCSPHIITHDMKSNVLLDENLEAIVSDPGARILMSAMDTHL	1041
Db	667	TRMKIAQDLAAGLFCILH--SOENIIHGNLTSSNVLLDENITNAKIALDFGLSKRLMSTRA-ANS	723
Qy	1042	SVSTLAGTPGVVPEPYXOSFRCSTKGVDVSYGVVLLTLLTGKRPDSDPFGNNILGWVK	1101
Db	724	NVIAVAGALGRAPRLSLKTKAANKTDTIYSGLVILLELLTRKSP--GVPNGGLDLPQWVA	781
Qy	1102	QHAKR-ISDVPDELMKEDALELIELLOHLKVAVACLDRANKRRTTMQVAMAFKEIJO	1159
Db	782	SVVKEEMNEVEFADDLMRDASTGVBDELLINTILKALHCVDPSPSGAPBEVHQVQOLEIR	840

```

: Sequence 1113, Application US/09754853A
: Publication No. US20030005491A1
: GENERAL INFORMATION:
:   APPLICANT: Hauge, Brian M.
:   APPLICANT: Parnell, Laurence D.
:   APPLICANT: Parsons, Jeremy D.
:   APPLICANT: Wang, Ming Li
:   TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
:   FILE REFERENCE: 38-10(15810)B
:   CURRENT APPLICATION NUMBER: US/09/754,853A
:   CURRENT FILING DATE: 2001-01-05
:   PRIOR APPLICATION NUMBER: US 60/174,880
:   PRIOR FILING DATE: 2000-01-07
:   NUMBER OF SEQ ID NOS: 1119
:   SEQ ID NO 1113
:   LENGTH: 854
:   TYPE: PRT
:   ORGANISM: Glycine max
:   FEATURE:
:   OTHER INFORMATION: Clone ID: rhg1_lee_amplicon
:   US-09-754-853A-1113
:
: Query Match      15.2%; Score 937; DB 9; Length 854;
: Best Local Similarity 31.3%; Pred. No. 1,7e-52;
: Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24;

```

```

QY 250 LKQOYLAMENKRTGEPDPLSGAOOTLGLDLSGNHFYGAVPFPGSCSLLESJLALS 349
Db 139 LOGRKRSLHDNQIGSSIPSTL-GLFENLKGVOFNNRLTGSJPLSGCPPLQSGJDLSEN 197
QY 350 NNFSGELPMDTLTKMRGLKAYLDLSFNEFSGELPESJLTNLASJLTLTDLSSNPFILPN 409
Db 198 NLRLGALPY-SLANSTKLYWNLINSFNSFGSPLPASJLTH-SFSLTFLSLQNNLSJGLSPNS 255
QY 410 LCQNPKN---TLQELYIQQNGFTGKIPPTLSNCSSELYSLHSFNYLSGTPRSSJGLSLKL 466
Db 256 WQGSNKGPFPLQWJLIDNPFETGSDVASJSGSREJNELJSLSNKTSFGALPNJGTLISRI 315
QY 467 RDLKJLMLNMEGEI-POELMWKTLTETJLIDFNDLTGEIPSGLSNCTNLWJLSNNRLTG 526
Db 316 K-----TLDISNNALNGJLPATLINSLSSTJLMLNENJLNLN 351
QY 527 EIPKWIQLENLIAJLKLSSNNSFSGNIPDEJGDCRSJLWJDLNLNLNFGTJPAAMFKQSG 586
Db 352 QIPSGJGLRMLSVJLISBRNQFSGHIPSJIANISJLSJQJDLSSNPFGEIPV----- 403
QY 587 IANANFIAGKRYVYIKNDGKKCKEAGANLLEFGJIRSEQLNRLSTNPNCTISRYVGGHT 646
Db 404 ----- 403
QY 647 SPTFDNNGSNMFLMSYNNLSGVIPEKIGSMPYJLJLNLGHNDJSGSIDPEVGDLRGJLNI 706
Db 404 --SFSQCSJLMLFNVSVXJSLSGSVP-----PLL----- 429
QY 707 LDLSSNKLJGRIPOAMSHJLWJLEJDLSSNNLSGPIPEMGQFETFPFAKJLNNPGLGCV- 765
Db 430 -----AKFNSSSFVGNJIOLOGYS 448
QY 766 PLPRG-----DPSNADGVAHHORSHGRPPASJAGSVAMGJLFSVCJFGLJLVG 814
Db 449 PSTPCLSQAPSGCVIAPPBVSJKHHHRKJLSTKOJLIIYAGVJLLVJLIIJLCCVJLFLCI- 507
QY 815 REMRKRKRKKAEJLEMYAEHGSNGD-----RTANNTNKKJLTVGXKA--LSJNLAAFE 865
Db 508 ---RKRJSJK-----AGNGQATEBGAAJMRTEKGVPPVLAAGDVABGAGAGKJLVHFD 556
QY 866 KPLRKLTPADJLJQATNGFHNDSLJGSGGFQDGYAYJLTKQGSAYAJKJLHYVSGQDREEM 925
Db 557 GPM-AFTADJLJCAT---AEIMKSYTGTVYFAJLJEDSGOVAVKJLRKJLREKJLTKGHEFE 610
QY 926 AMEJLITIGIKXKRNJLVJLLGYCKVG--DEJRLJVNEMVXGJSLJEDVJQDPKKGGYK--JLKS 981

```



Db 611 SEVSVLGKIRHNPVLAALRAY-YLGPKEKLLVPDYMKSGLASTLHG---CGTETPIDWP 666  
Qy 982 TRRKIAISGARGLAFLHNNCSPHIIHRDMKSNVLDENLEARYSDFGMARLMSAMDTLH 1041  
Db 667 TRMKIAODLARGLPCLH--SQENIIHGNTSSNVLDENTNAKADGSLKMTSTA-ANS 723  
Qy 1042 SVSTLAGTPGVVPEYQSFRCSTKGDVYSYGVVLELTGKRPDSDPDGNNLVGMVK 1101  
Db 724 NVIATAGALGYRABELSKLKANKTKTDIYSGLVILELTTRKSP--GVPMNGDLDPQWVA 781  
Qy 1102 QHAKLR-ISDVPDELMKEDPALIEILLOHLKVAACLDDBAMRPRPTMVOYMAFKEIQ 1159  
Db 782 SVVKEEMTNEVFADLMRDSSTVGDELLNTLTKALHCVDPSPARPRVHQVLOQLEIR 840

## RESULT 8

US-09-754-853A-1115  
; Sequence 1115, Application US/09754853A  
; Publication No. US20030005491A1  
; GENERAL INFORMATION:  
; APPLICANT: Hauge, Brian M.  
; APPLICANT: Parnell, Laurence D.  
; APPLICANT: Parsons, Jeremy D.  
; APPLICANT: Wang, Ming Li  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance  
; FILE REFERENCE: 38-10(15810)B  
; CURRENT APPLICATION NUMBER: US/09/754,853A  
; PRIOR FILING DATE: 2001-01-05  
; PRIOR APPLICATION NUMBER: US 60/174,880  
; PRIOR FILING DATE: 2000-01-07  
; NUMBER OF SEQ ID NOS: 1119  
; SEQ ID NO 1115  
; LENGTH: 854  
; TYPE: PRF  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: rhg1\_pi200499\_amplicon  
US-09-754-853A-1115

## Query Match 15.2%; Score 937; DB 9; Length 854;

Best Local Similarity 31.3%; Pred. No. 1,7e-52;  
Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24;

Qy 290 LKSLQVLSAENKFTGEIPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSS 349  
Db 139 LQGLRKLSLHDNIGSIPSTL-GLLPNLRGVQLFNNRLTGSIPSLGFCPLQSDLSN 197  
Db 350 NNFSGELPMDTLKMGKLVLDLSFNEFSGELPESLTNLSASLTLDLSSNNFSGPILPN 409  
Db 198 NLITGAIY-SLANSTKLVWLNSTFNSFSGPLASLTH-SFSLTFLSLONNMLSGSLPNS 255  
Qy 410 LCONPNK---TLQELYLONNGFTGKIPTLNSCSELVSLHLSFNYSGLTIPESLSLSKL 466  
Db 256 WGNSSKNGPFRLQNLILDNFFGDPVPSLSLRLEINISLSHNKFSGAIPIWEITLSRL 315  
Qy 467 RDLKLVANLLEGGIPELAMYVLTETLIDPDVLTGEISGSLNCTNMLNWLSSNNRLTG 526  
Db 316 K-----TLDISNNAALNGNLPATLSNLSLTLLAENNLIDN 351  
Qy 527 EIPKWLGRLENLAILKLSNNSFSGNIPDELGDCRSIIMDLNTNLFNNGIIPAAFKQSGK 586  
Db 352 QIPQISIGRLRNLSVILSNQFSGHIPSIANISLSLRQDLSLNNFSGEIPV----- 403  
Qy 587 IAAPIAGKRYYYIKNDKKECHAGNLEFQGISRSQNLNLSLRPNCTTSRYVGHT 646  
Db 404 ----- 403  
Qy 647 SPTFNNGSMFLDMSYNNLSGYIPKEIGSMXYLFLNLGNHNDIGSLPDEVGLRGINT 706  
Db 404 ---SPTDSGRNLFNYSNLSGSP-----PL----- 429  
Qy 707 LDISNKLIDGRIPOMSALVMTLEIDLSNNNLGPIPEWGFETETPPAKFLNPNGLCGY- 765

Db 430 -----AKFNSSPFQNIOLCGYS 448  
Qy 766 PLPRC-----DSANDVAHQHRRPRLASLAVMGLFSPVCFGLIIVG 814  
Db 449 PSTPLCSQAPSGVIAPEPEVSKHHHRKLSGTQDIIILVAGLVVLIILCVLLFCGLI- 507  
Qy 815 REMKRRRRKEAELEMYAEHGNSGD-----RPNANTNMKLITGVKEA---LSINLAPE 865  
Db 508 ---KRRSTK-----AGNQATEGPAATWRTKGVPPVAAGDVERGAGGKLVHFD 556  
Qy 866 KPLRLKTFADILLQATNGHNDLSLGGSGFVYKAILKDGSAVAIKKLIVHSGODRREFM 925  
Db 557 GPM-AFTADLLCAT-----AEIMKSTYGVYKAILLEDGQVAVKRLREKITGHEFE 610  
Qy 926 AEMETIGIKIRHNPVLAALRAY-YLGPKEKLLVPDYMKSGLASTLHG---CGTETPIDWP 666  
Db 611 SEVSVLGKIRHNPVLAALRAY-YLGPKEKLLVPDYMKSGLASTLHG---CGTETPIDWP 666  
Qy 982 TRRKIAISGARGLAFLHNNCSPHIIHRDMKSNVLDENLEARYSDFGMARLMSAMDTLH 1041  
Db 667 TRMKIAODLARGLPCLH--SQENIIHGNTSSNVLDENTNAKADGSLKMTSTA-ANS 723  
Qy 1042 SVSTLAGTPGVVPEYQSFRCSTKGDVYSYGVVLELTGKRPDSDPDGNNLVGMVK 1101  
Db 724 NVIATAGALGYRABELSKLKANKTKTDIYSGLVILELTTRKSP--GVPMNGDLDPQWVA 781  
Qy 1102 QHAKLR-ISDVPDELMKEDPALIEILLOHLKVAACLDDBAMRPRPTMVOYMAFKEIQ 1159  
Db 782 SVVKEEMTNEVFADLMRDSSTVGDELLNTLTKALHCVDPSPARPRVHQVLOQLEIR 840

## RESULT 9

US-09-754-853A-1110  
; Sequence 1110, Application US/09754853A  
; Publication No. US20030005491A1  
; GENERAL INFORMATION:  
; APPLICANT: Hauge, Brian M.  
; APPLICANT: Parnell, Laurence D.  
; APPLICANT: Parsons, Jeremy D.  
; APPLICANT: Wang, Ming Li  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance  
; FILE REFERENCE: 38-10(15810)B  
; CURRENT APPLICATION NUMBER: US/09/754,853A  
; PRIOR FILING DATE: 2001-01-05  
; PRIOR APPLICATION NUMBER: US 60/174,880  
; PRIOR FILING DATE: 2000-01-07  
; NUMBER OF SEQ ID NOS: 1119  
; SEQ ID NO 1110  
; LENGTH: 877  
; TYPE: PRF  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: rhg1\_noir\_amplicon  
US-09-754-853A-1110

## Query Match 15.2%; Score 937; DB 9; Length 877;

Best Local Similarity 31.3%; Pred. No. 1,8e-52;  
Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24;

Qy 290 LKSLQVLSAENKFTGEIPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSS 349  
Db 162 LQGLRKLSLHDNIGSIPSTL-GLLPNLRGVQLFNNRLTGSIPSLGFCPLQSDLSN 220  
Qy 350 NNFSGELPMDTLKMGKLVLDLSFNEFSGELPESLTNLSASLTLDLSSNNFSGPILPN 409  
Db 221 NLITGAIY-SLANSTKLVWLNSTFNSFSGPLASLTH-SFSLTFLSLONNMLSGSLPNS 278  
Qy 410 LCONPNK---TLQELYLONNGFTGKIPTLNSCSELVSLHLSFNYSGLTIPESLSLSKL 466  
Db 279 WGNSSKNGPFRLQNLILDNFFGDPVPSLSLRLEINISLSHNKFSGAIPIWEITLSRL 338



/ APPLICANT: Parsons, Jeremy D.  
 / APPLICANT: Wang, Ming Li  
 / TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 / FILE REFERENCE: 38-10(15810) B  
 / CURRENT APPLICATION NUMBER: US/09/754,853A  
 / PRIOR FILING DATE: 2001-01-05  
 / PRIOR APPLICATION NUMBER: US 60/174,880  
 / NUMBER OF SEQ ID NOS: 1119  
 / SEQ ID NO 1114  
 / LENGTH: 877  
 / TYPE: PRT  
 / ORGANISM: Glycine max  
 / FEATURE:  
 / OTHER INFORMATION: Clone ID: rhg1\_p1200499\_amplicon  
 / US-09-754-853A-1114

Query Match 15.2%; Score 937; DB 9; Length 877;  
 Best Local Similarity 31.3%; Pred. No. 1.8e-52;

Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24;

QY 290 LKSIQYLSAENKFTGEPDPLSGACDTLTGLDSGNHFGAVPFFGSCSLLESALSS 349  
 Db 162 LQGLRKLSDHNOIGSGIPSTL-GILPNLRGVOLFNNRLTSGIPSLGFCPLQSLDLSN 220  
 QY 350 NNFSGELPMDTLKRGKLVLDLSENFSGELPESLTNLSASLTTLDSNNFSGILPN 409  
 Db 221 NLTLGALPY-SLANSTLYLWNLSENFSGELPESLTNLSASLTTLDSNNFSGILPN 278  
 QY 410 LCONPKN--TLQELYLONNGFTGKIPTLSNCELSVLSHLSFNYLSGTIPESLSLSTL 466  
 Db 279 WGSNSKNKGFRLQNLIDNFFTGVDVPSLSRLNEISLSHNKSGALPNEIGTSLRL 338  
 QY 467 RDLKLMNLMBEIPQELMYVKTLETLIDPNDLTGEIPSGLSNCTNLMWISLSNNRLTG 526  
 Db 339 K-----TLDISNNALNGNLPATLSNLSLTLLAENMLDN 374  
 QY 527 EIPKWLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 Db 375 QIPQSLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 QY 587 IAAANFAGKRYVYIKNDGMKKECHAGNLLFQGISBQNLRLSTRNPNCTTSRYVGHT 646  
 Db 427 ----- 426  
 QY 647 SPTFDNNGSNMFLDMSYMLSGYIPKEISMPYLTILNGLNDISGIPDEVGLRGANI 706  
 Db 427 --SFDQSRSLNLFVSYNSLSGVP--PLL----- 452  
 QY 707 LDISNKLDRIPQAMSAITMLTEIDLSNNNSGPIPEMGQEPETPPAKFLNPGLCGY- 765  
 Db 453 ----- 426  
 QY 766 PLPRC-----DPSNADGYAHORSHGRPASLAGVAMGLLFSFVCIIGLIVG 471  
 Db 472 PSTPCLSQSPGCVTAPPEVSKHHHRKLTSTKDIILIVAGVLVYVILLCVLLPCLT- 530  
 QY 815 REMRKRRKKEAELEMYAECHGNSGD-----RTANNTNMGLTGKXA--LSLNLAAFE 865  
 Db 531 ---RKRSTSK-----AGNGQATEGPAATMRTKEGVPPVAAADVABGAGKLVHFD 579  
 QY 866 KPLRLKLPDILLQATNGSHNDLSIGSGFGDYVYAILKDGSAVAIKKLIHVSQGDREPM 925  
 Db 580 GPM-AFTADDDLCAT-----AEIMGKSTYGTIVYKALLLEDGQVAVKRLRKRTKGRHEFE 633  
 QY 926 AEMETIGIKIKRNLVPLLGYCKVG--DERLLINVMKYGSLIEDVLDPPKGGYK--LKLS 981  
 Db 634 SEVSUVTGIRHPNTALAAV-YLQPKGKELVDPYMSKGSILASFLNG--GGTETPIDMP 689  
 QY 982 TRRKAIASAGLAFLLHNCSPHIIRDMKSSNVLLDENTLFAVSDFGMARLSANDTHL 1041  
 Db 690 TRMKIAQDLARGLPCLH--SQENIIHGNTLSSNVLLDENTNAKIDAGLSRLMSTA-ANS 746

QY 1042 SVSTIAGTPGVPEVYQSFRCSTKGDVSYGVVLTLLTGKRPDSDPGDNNLYGWYK 1101  
 Db 747 NVIATVAGALGYRABELSKTKKXAKTKTDIYSLGVLLELTRKSP--GVMNGGLDQWYA 804  
 QY 1102 QHAKLR-LSVYFDELKPKEDPALEIELLQKLVAVACDDRAMRRRTMOWAMFKEIQ 1159  
 Db 805 SVVKEEWINEVFDADLRKDSIVYGDDELNTLKLALCHVPSPEARPEVHQVQLLEIR 863

## RESULT 12

US-09-754-853A-1098  
 / Sequence 1098, Application US/09754853A  
 / Publication No. US00030005491A1  
 / GENERAL INFORMATION:

/ APPLICANT: Hauge, Brian M.  
 / APPLICANT: Parnell, Laurence D.  
 / APPLICANT: Parsons, Jeremy D.  
 / APPLICANT: Wang, Ming Li  
 / TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 / FILE REFERENCE: 38-10(15810) B  
 / CURRENT APPLICATION NUMBER: US/09/754,853A  
 / PRIOR FILING DATE: 2001-01-05  
 / PRIOR APPLICATION NUMBER: US 60/174,880  
 / NUMBER OF SEQ ID NOS: 1119  
 / SEQ ID NO 1098  
 / LENGTH: 854  
 / TYPE: PRT  
 / ORGANISM: Glycine max  
 / FEATURE:  
 / OTHER INFORMATION: Clone ID: 240017\_region\_G3  
 / US-09-754-853A-1098

Query Match 15.1%; Score 933; DB 9; Length 854;  
 Best Local Similarity 31.4%; Pred. No. 3.1e-52;

Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;

QY 290 LKSIQYLSAENKFTGEPDPLSGACDTLTGLDSGNHFGAVPFFGSCSLLESALSS 349  
 Db 139 LQGLRKLSDHNOIGSGIPSTL-GILPNLRGVOLFNNRLTSGIPSLGFCPLQSLDLSN 197  
 QY 350 NNFSGELPMDTLKRGKLVLDLSENFSGELPESLTNLSASLTTLDSNNFSGILPN 409  
 Db 198 NLTLGALPY-SLANSTLYLWNLSENFSGELPESLTNLSASLTTLDSNNFSGILPN 255  
 QY 410 LCONPKN--TLQELYLONNGFTGKIPTLSNCELSVLSHLSFNYLSGTIPESLSLSTL 466  
 Db 256 WGSNSKNKGFRLQNLIDNFFTGVDVPSLSRLNEISLSHNKSGALPNEIGTSLRL 315  
 QY 467 RDLKLMNLMBEIPQELMYVKTLETLIDPNDLTGEIPSGLSNCTNLMWISLSNNRLTG 526  
 Db 316 K-----TLDISNNALNGNLPATLSNLSLTLLAENMLDN 351  
 QY 527 EIPKWLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 Db 352 QIPQSLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 QY 587 IAAANFAGKRYVYIKNDGMKKECHAGNLLFQGISBQNLRLSTRNPNCTTSRYVGHT 646  
 Db 404 ----- 403  
 QY 647 SPTFDNNGSNMFLDMSYMLSGYIPKEISMPYLTILNGLNDISGIPDEVGLRGANI 706  
 Db 404 --SFDQSRSLNLFVSYNSLSGVP--PLL----- 429  
 QY 707 LDISNKLDRIPQAMSAITMLTEIDLSNNNSGPIPEMGQEPETPPAKFLNPGLCGY- 765  
 Db 430 ----- 429  
 QY 766 PLPRC-----DPSNADGYAHORSHGRPASLAGVAMGLLFSFVCIIGLIVG 814

Db 449 PSTCLSQAPSGVIAPPEVSKHHHRKLTOKIILIVAGVLLVLLIICCVLLFCLT- 507  
 Qy 815 REMKRRKKEAELEMYAEGHNSGD-----RTANNTMVKLTGVKEA---LSINLAFE 865  
 Db 508 ---RKSTSK-----AGNGATEGRAATMTEKGVPPVAGGVVEAGGAGGLVHFD 556  
 Qy 866 KPLRLKLPADLLQATNGFHNDSLIGSGFGDYKAILKDGSAVAIKLHVSGQDREFM 925  
 Db 557 GPM-AFTADLLCAT-----AETMKSSTGYVYKAILLEDGSOVAVRLREKIKGHRFE 610  
 Qy 926 AEMETIGIKIRNVLPLGYCKVG--DERLLVNEVMKYSLEVLQDPKKGVC--LKLS 981  
 Db 611 SEVSVYGIKIRHPNVLLARAY-YLGPKEKLLVPDYMSKSLASFHNG--GGETFTDMP 666  
 Qy 982 TRKRIASARGAFLHNCSPHIIHRDMKSNVLIDENILEARVSDGMARLMSAMTHL 1041  
 Db 667 TRKRIADLARGLFCLH--SQENIIGHNLTSNVLLDENTNAKADGLSLMSTA-ANS 723  
 Qy 1042 SVSTLAGTPGVPEYOSFRCSTKGDVYSYGVVLELLTGKRPDTPDGDNNLVGVK 1101  
 Db 724 NVIATAGALGRAPBELSKLKANTKTIDYSLGVLLELLTRKSPGVSMNGLD--LPQWVA 781  
 Qy 1102 QHAKLR-ISDVPELMKEDEPALEIELLOHLKVAVACLDPMARPRPTMVMANFKEIQ 1159  
 Db 782 SVYKEWNTNEVFADLMRDASTVGDDELNTLKALHCVDPSPARPEVHQVLOQLEIR 840

RESULT 13  
 US-09-754-853A-1101  
 ; Sequence 1101, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; APPLICANT: Parsons, Jeremy D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754, 853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIOR APPLICATION NUMBER: US 60/174,880  
 ; PRIOR FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 1101  
 ; LENGTH: 854  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: rhg1\_A3244\_amplicon  
 US-09-754-853A-1101

Query Match 15.1%; Score 933; DB 9; Length 854;  
 Best Local Similarity 31.4%; Pred. No. 3.1e-52;  
 Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;  
 Qy 290 KLSLOYLSAEKFTBEIDFLSGACDTLTGDISGNHFGVAVPPFGSCSLLESIALS 349  
 Db 139 LOGRLKLSLHDQIGSISTLTGLIPNLRGVOLFNNRLTSGIPSLGCPGLQGLDLSN 197  
 Qy 350 NNFSGELPMDTLKMGKLVLDLSENFSGELPESTINLSASILLTDLSSNNSFGPILPN 409  
 Db 198 NLITGAIPY-SLANSRTKLVWLNLSFNSFGPLASLTH-SFSLTFSLSONNMISGLPNS 255  
 Qy 410 LCONPYN---TLQELYLONNGFGKIPPTLSNGSELVSLHSFNYSAGTIPSLGSLKL 466  
 Db 256 WGNSSKNGFRLQNLILDNHFFGADVPAISGSLRELNESLSNKSAGALPEIGLSTL 315  
 Qy 467 RDLKWLNMLEGEIPELWVVTLETLILDFNDLTGEIPSGLSNCTNLWISLNNRLTG 526  
 Db 316 K-----TLDISNNMLNGLPATTLSLSLTLNENNLIDN 351  
 Qy 527 EIPWIGRLNLAILKLSNNSFGNIPDELGDGRLIWLIDNTNLNGTIPAMFQSGK 586

Db 352 QIPQSLGRRLNLSVILSNNGSGHIPSISIANISLSRLQDLISLNNFSGEIPV----- 403  
 Qy 587 IANFTAGKRYIYIKNDGKKECHAGANLEFQGRSEQLNLSFRNCNITSRYVGGHT 646  
 Db 404 ----- 403  
 Qy 647 SPFDNNGSMWFLDMSYNNLSGYIPKEIGSMPEYFLNLGNDIGSLPDEVGLRGANI 706  
 Db 404 --SPDQSLNLFNVSYNLSGSVP-----PLL----- 429  
 Qy 707 LDLSNKLGRIPQMSALTMLTEIDLNNNLSGRIPEMGQFETPEPAKELNPGLCGY- 765  
 Db 430 -----AKFSSSFVGNIQCGIS 448  
 Qy 766 PLPRC-----DPSNADGYAHORSHGRPASLAGSVAMGLLFSVCIFGLIIVG 814  
 Db 449 PSTCLSQAPSGVIAPPEVSKHHHRKLTOKIILIVAGVLLVLLIICCVLLFCLT- 507  
 Qy 815 REMKRRKKEAELEMYAEGHNSGD-----RTANNTMVKLTGVKEA---LSINLAFE 865  
 Db 508 ---RKSTSK-----AGNGATEGRAATMTEKGVPPVAGGVVEAGGAGGLVHFD 556  
 Qy 866 KPLRLKLPADLLQATNGFHNDSLIGSGFGDYKAILKDGSAVAIKLHVSGQDREFM 925  
 Db 557 GPM-AFTADLLCAT-----AETMKSSTGYVYKAILLEDGSOVAVRLREKIKGHRFE 610  
 Qy 926 AEMETIGIKIRNVLPLGYCKVG--DERLLVNEVMKYSLEVLQDPKKGVC--LKLS 981  
 Db 611 SEVSVYGIKIRHPNVLLARAY-YLGPKEKLLVPDYMSKSLASFHNG--GGETFTDMP 666  
 Qy 982 TRKRIASARGAFLHNCSPHIIHRDMKSNVLIDENILEARVSDGMARLMSAMTHL 1041  
 Db 667 TRKRIADLARGLFCLH--SQENIIGHNLTSNVLLDENTNAKADGLSLMSTA-ANS 723  
 Qy 1042 SVSTLAGTPGVPEYOSFRCSTKGDVYSYGVVLELLTGKRPDTPDGDNNLVGVK 1101  
 Db 724 NVIATAGALGRAPBELSKLKANTKTIDYSLGVLLELLTRKSPGVSMNGLD--LPQWVA 781  
 Qy 1102 QHAKLR-ISDVPELMKEDEPALEIELLOHLKVAVACLDPMARPRPTMVMANFKEIQ 1159  
 Db 782 SVYKEWNTNEVFADLMRDASTVGDDELNTLKALHCVDPSPARPEVHQVLOQLEIR 840

RESULT 14  
 US-09-754-853A-1103  
 ; Sequence 1103, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; APPLICANT: Parsons, Jeremy D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754, 853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIOR APPLICATION NUMBER: US 60/174,880  
 ; PRIOR FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 1103  
 ; LENGTH: 854  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: rhg1\_peeking\_amplicon  
 US-09-754-853A-1103

Query Match 15.1%; Score 933; DB 9; Length 854;  
 Best Local Similarity 31.4%; Pred. No. 3.1e-52;  
 Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;

```

QY 290 LKSLQVLSAENKFTGEIPDLISGACDTLTGLDLSGNHYGAVPPFFGSCSLLESIALSS 349
DB 139 LOGLRKLSLHDNIGSISIPSTL-GILPNLRGVQLFNNRLTGSIPLSIGFPCPLQSLDLSN 197
QY 350 NNFSGELPMDTLTKMRGLKVLDSLSPNEFSGELPESLTNLSASLTTLDSNNNSGPILEN 409
DB 198 NLITGALPY-SLANSTKLYWNLSPNSFGPLPASLTH-SFSLTFSLQNNNLSSGLSPNS 255
QY 410 LCONPKN---TLQELYLONNGFTGKIPTLSNCSLVLHLSFNYLSGTPSSLSIGLSKL 466
DB 256 WGSNSKNGFPRLONLIDNHFPTGVPASLSGLRELNEISLSHNKFSGALPNEIGTSLRL 315
QY 467 RDLKMLNMLEGEIPQELMYVYLTLETLIDPNDLGEIPSGLSNCTNMTLSNNRLTG 526
DB 316 K-----TLDISNNALNGNLPATLSLSTLTLLNAENNLIDN 351
QY 527 EIPKWTGLENLAITKLNNSPSGNIPDELDCSLIWLIDNTNLFNGTTPAMFKQSGK 586
DB 352 QIPQSLGRRLNLSVILSRNQSCHIPSISLANISLSRLQDLSNNNSGELIPV----- 403
QY 587 IANFIAGKRYVYIKNDGKKECHGAGNLEFGQINSEQLNRLSTNPNCTITSRYVGHT 646
DB 404 ----- 403
QY 647 SPFDNNGSMWFLDMSYMLSGYIPKEIGSMPLYFLINLGHNDISGSIIDVGDRLGLNT 706
DB 404 --SPDSQSLNLFNVYSNLSGSP-----PLL----- 429
QY 707 LDISSNKLDGRIPQMSALTMLTEIDLSNNNLGSPIPMGQETFPPEAKFANNPGICGY- 765
DB 430 -----AKFNSSSFVGNIQCGYS 448
QY 766 PLPRC-----DPSNADGYAHHQSHGRRPASLAGSVAMGLFSFVCIFGLIIVG 814
DB 449 PSTPCLSQAPSGVIAPRPPEVSKHHHRKLSKTDIILYAGVLLVLIICCVLLFCLL- 507
QY 815 REMRKRKKKAELIEMAYABGHGNSGD-----RTANNWKLTGVKEA---LSINLAPE 865
DB 508 ---RKRSTSK-----AGNGQATEGRAATMTEKGVPPVAGADVEAGGAGKLVHFD 556
QY 866 KPLRKLTFADLLOATNGFHNDSLIGSGFGDYKAILKQGSAAVAKLIVHSGQDREEM 925
DB 557 GPM-AFTADDLICAT-----AETMGKSTYGTIVYKALIEGSOVAVRLEKTIKGHREFE 610
QY 926 AEWETIGIKIRNLVPLGCKVG--DERLLVNEWVKYGSLEVDVQDPKKGCVK--LKLS 981
DB 611 SEVSVLGKIRHPVNLALRAY-YLGPKEKLLVFDYMSKGSLSAFLHG--GATEFTIDWP 666
QY 982 TRRKIIAGSARGLAFLHNHCSPHIIRDMKSSNVLLDENLEARVDFGMARLMSAMDTHL 1041
DB 667 TRMKIIADLARGLFCLH--SQENIIHGNTLSSNVLLDENTYAKIADFGSLRIMSTA-ANS 723
QY 1042 SVSTIAGTGPVPEPYGFRCTGKDYVSYGVVLELLTGKRPDPSDFGDNMLVGMYK 1101
DB 724 NVIATGALGYRAPBELSKLKANTKTIDYSLGVILLELLTRKSPGVSNMGD--LPQWVA 781
QY 1102 QHAKLR-ISDVPDEIMKEDPALIEILLKVAACLDDBAMRPTMVMAMFKIQ 1159
DB 782 SVYKEEMTNEVFDADLMDASTVGBELLNTIKALHCVDPSPSAKPEVHOVLQGLEIR 840

```

RESULT 15  
US-09-754-853A-1105

; Sequence 1105, Application US/09754853A  
; Publication No. US20030005491A1

; GENERAL INFORMATION:

; APPLICANT: Hauge, Brian M.

; APPLICANT: Patneil, Laurence D.

; APPLICANT: Parsons, Jeremy D.

; APPLICANT: Wang, Ming Li

; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With

; FILE REFERENCE: 38-10(115810)B

```

; CURRENT APPLICATION NUMBER: US/09/754, 853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174, 880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1105
; LENGTH: 854
; TYPE: PR
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_toyosuzu_amllicon
US-09-754-853A-1105

Query Match      15.1%; Score 933; DB 9; Length 854;
Best Local Similarity 31.4%; Pred. No. 3,1e-52;
Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;

QY 290 LKSLQVLSAENKFTGEIPDLISGACDTLTGLDLSGNHYGAVPPFFGSCSLLESIALSS 349
DB 139 LOGLRKLSLHDNIGSISIPSTL-GILPNLRGVQLFNNRLTGSIPLSIGFPCPLQSLDLSN 197
QY 350 NNFSGELPMDTLTKMRGLKVLDSLSPNEFSGELPESLTNLSASLTTLDSNNNSGPILEN 409
DB 198 NLITGALPY-SLANSTKLYWNLSPNSFGPLPASLTH-SFSLTFSLQNNNLSSGLSPNS 255
QY 410 LCONPKN---TLQELYLONNGFTGKIPTLSNCSLVLHLSFNYLSGTPSSLSIGLSKL 466
DB 256 WGSNSKNGFPRLONLIDNHFPTGVPASLSGLRELNEISLSHNKFSGALPNEIGTSLRL 315
QY 467 RDLKMLNMLEGEIPQELMYVYLTLETLIDPNDLGEIPSGLSNCTNMTLSNNRLTG 526
DB 316 K-----TLDISNNALNGNLPATLSLSTLTLLNAENNLIDN 351
QY 527 EIPKWTGLENLAITKLNNSPSGNIPDELDCSLIWLIDNTNLFNGTTPAMFKQSGK 586
DB 352 QIPQSLGRRLNLSVILSRNQSCHIPSISLANISLSRLQDLSNNNSGELIPV----- 403
QY 587 IANFIAGKRYVYIKNDGKKECHGAGNLEFGQIRISQRLNRLSTNPNCTITSRYVGHT 646
DB 404 ----- 403
QY 647 SPFDNNGSMWFLDMSYMLSGYIPKEIGSMPLYFLINLGHNDISGSIIDVGDRLGLNT 706
DB 404 --SPDSQSLNLFNVYSNLSGSP-----PLL----- 429
QY 707 LDISSNKLDGRIPQMSALTMLTEIDLSNNNLGSPIPMGQETFPPEAKFANNPGICGY- 765
DB 430 -----AKFNSSSFVGNIQCGYS 448
QY 766 PLPRC-----DPSNADGYAHHQSHGRRPASLAGSVAMGLFSFVCIFGLIIVG 814
DB 449 PSTPCLSQAPSGVIAPRPPEVSKHHHRKLSKTDIILYAGVLLVLIICCVLLFCLL- 507
QY 815 REMRKRKKKAELIEMAYABGHGNSGD-----RTANNWKLTGVKEA---LSINLAPE 865
DB 508 ---RKRSTSK-----AGNGQATEGRAATMTEKGVPPVAGADVEAGGAGKLVHFD 556
QY 866 KPLRKLTFADLLOATNGFHNDSLIGSGFGDYKAILKQGSAAVAKLIVHSGQDREEM 925
DB 557 GPM-AFTADDLICAT-----AETMGKSTYGTIVYKALIEGSOVAVRLEKTIKGHREFE 610
QY 926 AEWETIGIKIRNLVPLGCKVG--DERLLVNEWVKYGSLEVDVQDPKKGCVK--LKLS 981
DB 611 SEVSVLGKIRHPVNLALRAY-YLGPKEKLLVFDYMSKGSLSAFLHG--GATEFTIDWP 666
QY 982 TRRKIIAGSARGLAFLHNHCSPHIIRDMKSSNVLLDENLEARVDFGMARLMSAMDTHL 1041
DB 667 TRMKIIADLARGLFCLH--SQENIIHGNTLSSNVLLDENTYAKIADFGSLRIMSTA-ANS 723
QY 1042 SVSTIAGTGPVPEPYGFRCTGKDYVSYGVVLELLTGKRPDPSDFGDNMLVGMYK 1101
DB 724 NVIATGALGYRAPBELSKLKANTKTIDYSLGVILLELLTRKSPGVSNMGD--LPQWVA 781

```

Wed Mar 12 10:45:02 2003

us-09-823-394-2.rapb

Page 12

QY 1102 QHAKLR-1SDVDPPEIMKEDPALTEILOHLKVAVACLDDBRAMRRPTWVOVWAMFKXIQ 1159  
DB 782 SVVKEEMTNEVPDADLMDASTVGDELLNTILKDLALHCVDPPSPARPEVHQVLOQLEIR 840

Search completed: March 10, 2003, 18:11:42  
Job time : 34.6693 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:03:39 / Search time 25.6693 Seconds

(without alignments)  
1370.890 Million cell updates/sec

Title: US-09-823-394-2

Perfect score: 6183

Sequence: 1 MKTSSFFLSTVTLTFPSF.....GFSTIEMWMSIKVEEGL 1196

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: /cgnt\_6/prodata/2/1aa/5A COMB pep: \*  
2: /cgnt\_6/prodata/2/1aa/5B COMB pep: \*  
3: /cgnt\_6/prodata/2/1aa/5C COMB pep: \*  
4: /cgnt\_6/prodata/2/1aa/5D COMB pep: \*  
5: /cgnt\_6/prodata/2/1aa/5E COMB pep: \*  
6: /cgnt\_6/prodata/2/1aa/5F COMB pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6183	100.0	1196	US-08-881-706-2	Sequence 2, Appl1
2	1311	21.2	999	US-08-473-553A-5	Sequence 5, Appl1
3	1225.5	19.8	980	US-08-473-553A-6	Sequence 6, Appl1
4	1218.5	19.7	985	US-08-473-553A-2	Sequence 2, Appl1
5	1119.5	18.1	910	US-09-228-986-72	Sequence 72, Appl1
6	1087	17.6	1025	US-08-587-375-4	Sequence 4, Appl1
7	1087	17.6	1025	US-08-587-375-4	Sequence 4, Appl1
8	1073.5	17.4	1012	US-08-475-881A-4	Sequence 4, Appl1
9	1057	17.1	947	US-09-228-986-73	Sequence 73, Appl1
10	891	14.4	1112	US-09-353-585-2	Sequence 2, Appl1
11	890	14.4	1016	US-09-180-439-8	Sequence 8, Appl1
12	889	14.4	1112	US-09-353-585-3	Sequence 3, Appl1
13	885	14.3	968	US-09-180-439-3	Sequence 3, Appl1
14	884	14.3	968	US-09-180-439-3	Sequence 3, Appl1
15	875.5	14.2	1023	US-08-475-881A-2	Sequence 2, Appl1
16	875.5	14.2	1023	US-08-475-881A-2	Sequence 2, Appl1
17	875.5	14.2	1023	US-08-567-375-2	Sequence 2, Appl1
18	875	14.2	1023	US-08-567-375-2	Sequence 2, Appl1
19	798	12.9	630	US-09-228-986-76	Sequence 76, Appl1
20	796	12.9	630	US-09-228-986-76	Sequence 76, Appl1
21	795.5	12.9	630	US-09-228-986-76	Sequence 76, Appl1
22	759	12.3	638	US-09-180-439-6	Sequence 25, Appl1
23	715.5	11.6	805	US-08-945-983-2	Sequence 74, Appl1
24	705.5	11.4	863	US-08-666-271-2	Sequence 2, Appl1
25	673	10.9	666	US-09-228-986-68	Sequence 68, Appl1
26	668.5	10.8	655	US-09-228-986-70	Sequence 70, Appl1
27	668.5	10.8	659	US-09-228-986-75	Sequence 75, Appl1

28	655.5	10.6	707	US-09-228-986-80	Sequence 80, Appl1
29	649.5	10.5	711	US-09-228-986-79	Sequence 79, Appl1
30	629.5	10.2	523	US-08-473-553A-3	Sequence 3, Appl1
31	589.5	9.5	903	US-09-228-986-78	Sequence 78, Appl1
32	575	9.3	690	US-09-228-986-69	Sequence 69, Appl1
33	524	8.5	858	US-08-265-628-2	Sequence 2, Appl1
34	508.5	8.2	282	US-08-473-553A-4	Sequence 4, Appl1
35	498.5	8.1	857	US-07-717-331E-3	Sequence 3, Appl1
36	497	8.0	857	US-07-717-331E-2	Sequence 2, Appl1
37	466	7.5	321	US-08-447-185-1	Sequence 1, Appl1
38	445	7.2	739	US-09-503-922-1	Sequence 1, Appl1
39	428	6.9	632	US-09-228-986-77	Sequence 77, Appl1
40	384.5	6.2	501	US-08-980-060-6	Sequence 6, Appl1
41	384.5	6.2	501	US-09-307-185-6	Sequence 6, Appl1
42	377	6.1	712	US-08-587-889-2	Sequence 2, Appl1
43	377	6.1	712	US-08-587-889-2	Sequence 2, Appl1
44	377	6.1	712	US-09-307-185-5	Sequence 5, Appl1
45	377	6.1	712	US-09-307-185-5	Sequence 5, Appl1
			5	PCT-US96-09193-2	Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-08-881-706-2  
Sequence 2, Application US/08881706  
Patent No. 6245969  
GENERAL INFORMATION:  
APPLICANT: Chong, Joane  
TITLE OF INVENTION: Receptor Kinase BIN1  
FILE REFERENCE: 07251/022001  
CURRENT APPLICATION NUMBER: US/08/881,706  
CURRENT FILING DATE: 1997-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1196  
TYPE: PRT  
ORGANISM: Arabidopsis sp.  
US-08-881-706-2

Query Match	Best Local Similarity	100.0%; Score 6183;	DB 4;	Length 1196;
Matches 1196;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKTSSFFLSTVTLTFPSFSLFQASPSQGLYREIHQLISFKDVLDPDKMLPDMSSNKN	60	
DB	1	MKTSSFFLSTVTLTFPSFSLFQASPSQGLYREIHQLISFKDVLDPDKMLPDMSSNKN	60	
QY	61	PCTPGVTCRDKKTSIDLSKSPINVGSAVSSLLSLTGLESFLNSHINGSVGFKC	120	
DB	61	PCTPGVTCRDKKTSIDLSKSPINVGSAVSSLLSLTGLESFLNSHINGSVGFKC	120	
QY	121	SASLTSIDLSNRSISGPVTLTSLGSCGKFLNVSNTLDPFGKVSGLKMSLVLLD	180	
DB	121	SASLTSIDLSNRSISGPVTLTSLGSCGKFLNVSNTLDPFGKVSGLKMSLVLLD	180	
QY	181	SANSTSGANVGMWLSGCCGKFLNVSNTLDPFGKVSGLKMSLVLLD	240	
DB	181	SANSTSGANVGMWLSGCCGKFLNVSNTLDPFGKVSGLKMSLVLLD	240	
QY	241	IGDSCALQHLDISGNKLSGDFSRISTCTELKLNLSNQFVPIPLKSLQYISTLAE	300	
DB	241	IGDSCALQHLDISGNKLSGDFSRISTCTELKLNLSNQFVPIPLKSLQYISTLAE	300	
QY	301	NKFTGELPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT	360	
DB	301	NKFTGELPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT	360	
QY	361	LKKRGKLVLDLSEFNEFSGELPESLTNLSASLTLTDLSSNNFSGELPMDT	420	
DB	361	LKKRGKLVLDLSEFNEFSGELPESLTNLSASLTLTDLSSNNFSGELPMDT	420	
		LKKRGKLVLDLSEFNEFSGELPESLTNLSASLTLTDLSSNNFSGELPMDT	420	

	MEDIUM TYPE:	Floppy disk
	COMPUTER:	IBM PC compatible
	OPERATING SYSTEM:	PC-DOS/MS-DOS
	SOFTWARE:	Patentin Release #1.0, Version #1.30
	CURRENT APPLICATION DATA:	
	APPLICATION NUMBER:	US/08/473,553A
	FILING DATE:	06-JUN-1995
	CLASSIFICATION:	800
	ATTORNEY/AGENT INFORMATION:	
	NAME:	Silva, Robin M.
	REGISTRATION NUMBER:	38,304
	REFERENCE/DOCKET NUMBER:	A-60886/RFT/RMS
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE:	(415) 781-1989
	TELEFAX:	(415) 398-3249
	TELEX:	910 277299
/	INFORMATION FOR SEQ ID NO:	5:
/	SEQUENCE CHARACTERISTICS:	
/	LENGTH:	999 amino acids
/	TYPE:	amino acid
/	STRANDEDNESS:	unknown
/	TOPOLOGY:	unknown
/	MOLECULE TYPE:	protein
/	US-08-473-553A-5	
Query Match	21.2%; Score 1311; DB 2;	Length 999;
Best Local Similarity	32.4%; Pred. No. 3,96-87;	
Matches	386; Conservative 161; Mismatches 392;	Indels 254; Gaps 30
Oy	22 LSPQASPSQSIRKRIHQSLSFKHVLPD-KULLPDWSSNKN--PCTPGVCYTC-RDDKYTSI	77
Dd	11 LSTYLPLSLNDQATTLROAKIGLSDPPQASSMSDNVDTPCKMIGVSCDATSNVSV	70
Oy	78 DLSKFLNWGSAVSSLSLTGLBESLFISNHINGSVSG-FKCSALITSLDLRSSLS	135
Dd	71 DLSSTMVGPPSI---LCHLPSLHSLSLYNNISINGSLADPDFCHNLISLISENLIV	127
Oy	136 GPVTTLTSLGSCSGLKFLNVSNITDPFGKVSGGLKNSLEVLDLSANSISGANVGVML	195
Dd	128 GSIP-----KSLEPIPL-----	139
Oy	196 SBOCGELHIALISGNKISGDVDVSRCVNLEFLDVSSNNFSTGI-PFLDCSALQHLDISG	254
Dd	140 -----NLKELEIGNNLSDTISSFGFPRLSELMLAG	172
Oy	255 NKLGGDFRAISTCTELKLNINISSNQF-VGPPIPL--PLKSYOVLSENKRTGBIPDL	311
Dd	173 NFLSGTIPASLGNTTKLKELKLAVALNFSPSOIPSOLGNTELQVIMLAGCNLVETIPSL	232
Oy	312 SGACDTLTGLDSGNHFYGAVPFPGSCSLLESIALSSNNFSGELPMDTLLMRGLAYLD	371
Dd	233 S-SLTVLVMIDLTFQLTGSSIPSWITQKTVEQJELFNNSFSGEIP-EEMGMVTTLKRPD	290
Oy	372 LSFRNESGELPESLTNLSASLITLTDSSNFFSGEIPENLCQNPKNLTLOELYQNNGCFPK	431
Dd	291 ASNMKATGKLPDNIL-NLESINLPFNMLEGPLPESTIRS--KTLSEIKLFNNRNLGV	346
Oy	432 IIPPTLNCSELVSLHSENYISGTIPSSLGLSLTRDKLMLMGEGEPDEIMVYKTL	491
Dd	347 LPQLGANSPLQYVDYSYNRFSGELPAVACGEGLLEYLLIDNSFSGEISNNLKGRKSLT	406
Oy	492 TLLIDPMDLTGELPSGLASCNTLMMNISNRNLGELPKMIGRELENLAIIKLSNSSFCA	551
Dd	407 RVRLSNNRKSGQIPHGFMGIPRLSLIEISDMSFPGSIPKTIIGAKNLSNRISRISKFSSGS	466
Oy	552 IPDELGCRRSLIMLDINTNLFNGTIPAMAFKOSGKIAPLANFAIGRYYIINDMKKECHG	611
Dd	467 IPNELGINSGLIISGANDFSGEIPESL-----	495
Oy	612 AGNILEFOGIRSEQNLRTKPNCNITSRYVGHTSPFDNNGSMFMFLDMSYNMLSGYIP	671
Dd	496 -----VTLKQLSR-----LDLSKQQLSEIP	516

```

QY 672 KEIGSWPYFLINLGHNDISGIPDEVDLGLNLIDLSNKLDRIPQAMSLTMLTEI 731
DB 517 RBLRCKNMLNELLANNHSGEIPKEVGLPLVLYLIDLSNQSGEIPLEQLN- KLVNL 575
QY 733 DLSNNLSGPIPEMGOFTFPKAPLANNPGLCGVPLPRCPSNADGVAHHQRSGRRPAS 791
DB 576 NLSVYHLSGKIPPLVANNITY-AHPFISGPGC-----VDLDGLCR-----K 615
QY 792 LAGSVAMG-----LFSFVCIPLGLVGRMRKRRKKEALEMYAAGHNSGDR 842
DB 616 ITRSNKIGYVILLTIFLAGIVFVGIYMFIAKCRKLKALKSTLLA----- 663
QY 843 ANNTMKLTGVKALSTINLAPEKRLKTF-----ADLLQATNGFNDLSISGGFGDY 897
DB 664 ---SKM-----RSFHLHSEHEIADCLDEKN-----VIGFGSGKV 697
QY 898 YKAILIKOGSAVAIKKLIVHSGQGDRE-----FMAEMETIGIKRNLVPLIGYCK 947
DB 698 YKVELRGEVAVAKKLNKSVKGGDEVSQSLNRDVAEVELTGTIRKSTIRLMCCS 757
QY 948 VGBRLVNEVMKGSLEEDYQDPKGVYKUKSTRKIAISGAGIAFLHNSCPHIIH 1007
DB 758 SGDCCLVETWPGSLADYLAHGRKGVVIGWPERLRIALDAEGLSYLHDCVPIVH 817
QY 1008 RDMKSNVLDENLEARVDFGMARL--MSAMDHLVSSTLACTPGVPEPYQSFRCST 1065
DB 818 RDVSSNILLDSQDGAVALDFIAKVGOMSGSKTPEMGSIGASCYIAPEVYVYTLRVNE 877
QY 1066 KGVYISGVVILELITGRPTDSPDFGNNLVGMV-KOHAKLRISVPPPELKEPDL 1124
DB 878 KSDIYSGVVLLELVTSKQPTDS-ELGDKDMAMVCTALDKGLEVIDPKL---DLRFK 933
QY 1125 IELLOHKVAVACLDDBAMRPTMVOVMFKELIQAGSGIDSGSTIRSIJEDG 1177
DB 934 ERSKVTIHGILCTSPPLNRPSPMRKVVIMLDEVGAVPCSSPNTSKSKTG 986

RESULT 3
US-08-473-553A-6
; Sequence 6, Application US/08473553A
; Patent No. 5859338
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliot M.
; APPLICANT: Clark, Steven E.
; APPLICANT: Williams, Robert W.
; TITLE OF INVENTION: Plant Clavatal Nucletic Acids,
; TITLE OF INVENTION: Transformed Plants, and Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Teet, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,553A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 980 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-473-553A-6

Query Match 19.8%; Score 1225.5; DB 2; Length 980;
Best Local Similarity 29.4%; Pred. No. 6.5e-81;
Matches 352; Conservative 189; Mismatches 356; Indels 297; Gaps 35;

QY 14 LFFSPFSLSFQASPSQSLYREIHOLISFKDYL--PDKRLDMSNKNP---CTPDGYT 68
DB 13 LHLVLFSPCF-----AYTMEVLLNKSSWIGPKGGLHDMTHSSPDHACSPSGVS 65
QY 69 CRDD-KVTSIDLSKPLNVPFSAVSSSLSLTGLTESLFTSNHINGSVGFRCASLSLSL 127
DB 66 CDDARVVISLNVSFPTPL--FGTISPEIGMLTHVNLTLAANNFTG-----EL 110
QY 128 DLSRNSLSGVTTLTSLGSCSGIKPLNVSSN--TLDPEKVSGLKLNLEVLDSLANS 184
DB 111 PLENKSL-----TSIKVNLINNNGLITGPEGEI----- 139
QY 185 ISGANVGVVLDGCGELKHLAISGNKISGDVDSRCVNLIELDVSSNNFSTGIPFLGDC 244
DB 140 -----LKAVDLEVLDTYNNFN----- 157
QY 245 SALQHLDISGNKISGDFSAISTCTELKLNINSSNQFVPIPP--LPKSLQYSLAENK 302
DB 158 -----GKLPEMSELKTKYLSFGGNF 179
QY 303 FTGEIPDPLSGACDNLTLGLDSGNHFGAVPFFGSCSLSLALS-SNNSFGEIPMD-- 359
DB 180 FSGELPEST-GRIOSELYIGLNGAGLSKSPAPFLSRLKNLEMYIGVYNSYTGVPREFG 238
QY 360 TILKRGKLVLDLSFNEPSGELPESLTNLSASLTLTDLSSNPFGLPLNLCQPKXTIQ 419
DB 239 GLTK--LEILMDASCTLTGLIPTLSNL-----KHLH 268
QY 420 ELYLQNGFTGKIPTLNSCELYSLHSFNYLSGTTIPSSLSLSKLDKLMMLMEGE 479
DB 269 TFLHNNMLTGHIPPELGLVSLKSLDISIQNGEIPQSPINLGNITLPLFNNLYGQ 328
QY 480 IPOELMYVYTLTETLIDPNDLTGEIIPSGLSNCTLNMISSNNRLTGEIKYGRLENLA 539
DB 329 IPEALGELPKLEVFVEWENFTLQIPAVLGRNGLIKLDVSDNHLTGLIPKDLCRGKLE 388
QY 540 IUKLSNNSPSGNIPELDGDCRLIMLDINTNLFNGTIPAMFK-----QSGKIANFAGK 595
DB 389 MLILSNNEFFGPIPELIGCKSLKIRIVKNLNGIVPAGLPLVTTIELDNPFSGE 448
QY 596 RYVYIKNDGMKKECHGAGNLEFQGISBQNLSTNPNCTISRYVGGTSPFPNNGS 655
DB 449 LPVWISGVL-----DQYILSNMFSGEIPPAIGNPN 481
QY 656 IMFILDSYMLSGYIPKEIGSPYFLINLGHNDISGIPDE----- 697
DB 482 LQTLUDLNNRFRGNIPELFEIKHLSKINTSANNITGIPDGSISRCSTLISVDSLNRIN 541
QY 698 -----VDLRLGLNLDLSNKLDRIPQAMSLTMLTEIDLSNNMISGPIPEMGOFTFP 751
DB 542 GEIPKGINNVKLTGLINISGQULGSIPGTIGMTSLTTLTDLSDFNLSGRVPLGGFPLVF 601
QY 752 PPAKPLNPGLCGVPLR---CDPSNADGYAHHQRSGRRPASLAGSVAMGLIFSVCIF 808
DB 602 NETSPAGNTYIC--LPHRYSC-PTRPGQTSDDNHTALSPSPSIVITV-----IAAIT 650
QY 809 GLIIVGRMRKRRKKEALEMYAAGHNSGDRTANNVWKLTGVKALSTINLAPEKRL 868
DB 651 GLILISVAIRQMKKKKQK-----SLAMKLT-----AFQ--- 679
QY 869 RKLTF--ADLLQATNGFNDLSISGGFGDYKAILIKOGSAVAIKKLIVHSGQGDREFM 925

```

Db 680 -KIDFSEVLEEC---LKENIIGKSGIYVRGSMNNDVAIKRLVGRGGRSDHGET 735  
 QY 926 AEMETIGIKRNLVPLLGCKYGERLLVNEVMKGSLEVDLPDPKGGVYLKXSTRK 985  
 Db 736 AEIQTGRIRHRIYVLLGVANKDNTLLIYEMPGSIGELHSGKG--HLQWETHR 793  
 QY 986 IAGSARGLAFLHNSGPHIHRDMKSSNVLDENLEAVSPFGMARLMSAMDTLSVST 1045  
 Db 794 VAVEAKGGLCYLHSDSPILHRDVSNMILLSDPEAHVADFGAKFLVDGAASECMS 853  
 QY 1046 LAGTPGVPEYQSPFCSTKGDVSYGVLLLELTGKRPTSPDFGDN-NLVGWYKQHA 1104  
 Db 854 IADSYGIAPEYAVYTLKVDKSDVSGVLLLELTGKRPTSPDFGDN-NLVGWYKQHA 910  
 QY 1105 KAIRSDVPELMEKE--DPALE---IELLOHLKVAVACLDRAKRPMTVOYMA 1154  
 Db 911 EEBITQSDAIVAVIADPRLTGYPILTSVHVFKIAMCVBEBAARPTMREVVM 966

LET 4

US-08-473-553A-2  
 ; Sequence 2, Application US/08473553A  
 ; Patent No. 5859338  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meyerowitz, Elliot M.  
 ; APPLICANT: Clark, Steven E.  
 ; APPLICANT: Williams, Robert W.  
 ; TITLE OF INVENTION: Plant Clavate Nucleic Acids,  
 ; TITLE OF INVENTION: Transformed Plants, and Proteins  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/473,553A  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Silva, Robin M.  
 ; REGISTRATION NUMBER: 38,304  
 ; REFERENCE/DOCKET NUMBER: A-060866/RFT/RMS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 781-1989  
 ; TELEFAX: (415) 398-3249  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO. 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 985 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-473-553A-2

Query Match 19.7%; Score 1218.5; DB 2; Length 985;  
 Best Local Similarity 29.3%; Pred. No. 2.1e-80;  
 Matches 351; Conservative 189; Mismatches 359; Indels 297; Gaps 35;

QY 14 LFFPSPSLSPQSPQSLYREIHOLISFQDVL--PDKNLLPQMSNKNP---CTPDGYT 68  
 Db 18 LHLFFSPSPCF-----ATTDEVLNLTSSKMTGKGGHLDHMHSSSPDACHSGSVGS 70  
 QY 69 CRDD-KVTSIDLSKPLNWFSAVSSLSLTGLTESLPLSNHSHNGSVSGFKCASLTSL 127

Db 71 CDDDAVIGLNVSPFTL---FGTISPEIGMLTHLVNLTLANNTG-----EL 115  
 QY 128 DLSRNSISGEVTTTLTSGSCGSLKPLNVSN---TLDFPGXVSGGLKNSLBYLDLSANS 184  
 Db 116 PLEMKSL-----TSLKVLNINSGNLGTFFGEI-----144  
 QY 185 ISGANVGVWVLDSCGELKHLAISGNKISGDVDSRCVNIIEFLDVSSNNFSGIPLGDC 244  
 Db 145 -----LKAMVLEVDLTJNNNNFN-----162  
 QY 245 SALQHLDISGNKLSGDFSRRAISTCTELKLNLSNQFVGPIPP--LPLKSLOYLSAENK 302  
 Db 163 -----GLPEMSELKKLTKLSTFGCNF 184  
 QY 303 FTGRIPEPLSGADDTLTGLDLSGNHFGAVPPFGSCSLSEIALS--SNFSGELPMD-- 359  
 Db 185 FSGEIPESY-GDIQSLLEYLUNGAGLSGKSPALSRUKLRMYIGVNSYTGVRREFG 243  
 QY 360 TLLKWRGLKVLDSFNEFSGELPESLTNLSASLLTLDSSNNFSGPIPLNLCOMPKNITQ 419  
 Db 244 GLTK---LEILDMACTLTGEIPTSLSNL-----KHLH 273  
 QY 420 ELYIQNNGFTGKTIPTLSNCSBELVSLHSFNYSCTIPSSLSGSLKDLKMLNMLEGE 479  
 Db 274 TLFPHINNLTHIPEPLSGVLSKSLDLSINQLTGEIPQSFINGITLTLNLRNNLYGQ 333  
 QY 480 IPEHLMVYKLTETLLDPNDLTGEIPSGLSNCTNLMNISNNRLTGEIPKMGRLNLA 539  
 Db 334 IPEALGELPKLEVEWENNTTLDLPANLGNGLIKLDVSDNHLTGLIPDLORGEKLE 393  
 QY 540 ILKLSNNSFSGNIDELGDCSLIWLNLNTNLNFGTIPAMFK-----QSGKLANFAGK 595  
 Db 394 MLISNNSFFGPIPELCKCKSLTKIRLVKNLNGVAPAGLNLPLVTIIELTNPFSGE 453  
 QY 536 RYVYIKNDKMKCKEAGANLEFGQIRSEQLNRLSTRNPTCVTSRYVGHTSPPTDNGS 655  
 Db 454 LPVTVSGDVL-----DQIYLSNNWSGELIPALGNFPN 486  
 QY 656 MMFLMSYNNLSGYIPKEIGSNPYLFIPLGNHDSISIDE-----697  
 Db 487 LQTLPLDRRFRGNIPREIPELKHLSRINTSANNITGIPDISISRCTSLTIVDSLRNRIN 546  
 QY 698 -----VGLRGELNLDLSNKLDRIPQMSALTMLEIDLSNNNSGPIPEMGOFFET 751  
 Db 547 GEIPKINNVMKLGTLNISGNQLTGSIPTGIGNMTSLTLDSEFVDSGRVPLGGQFLVF 606  
 QY 752 PPAKFLNPNGLGYPLR---CDPSNADGYAHQSHGRRRPAPSLAGVAMGLSFVCIF 808  
 Db 607 NETSPAGNTYLC---LPHRVSC-PTRPGQSDHNHTALFSPERIVIV-----IAAIT 655  
 QY 809 GLILVGRNKRKRRKKEALEMYAEGHNSGDRITANNNTMKLTGYKEALSTILAPEKPL 868  
 Db 656 GLILISVAIRNNKKKNQK-----SLAMKLT-----AFQ---684  
 QY 869 RKLTF--ADLLQATNGFNHDSLISGSGPDYKATIKDQSAVAIKLI-HVSGQDREEM 925  
 Db 685 -KIDFSEVLEEC---LKENIIGKSGIYVRGSMNNDVAIKRLVGRGGRSDHGET 740  
 QY 926 AEMETIGIKRNLVPLLGCKYGERLLVNEVMKGSLEVDLPDPKGGVYLKXSTRK 985  
 Db 741 AEIQTGRIRHRIYVLLGVANKDNTLLIYEMPGSIGELHSGKG--HLQWETHR 798  
 QY 986 IAGSARGLAFLHNSGPHIHRDMKSSNVLDENLEAVSPFGMARLMSAMDTLSVST 1045  
 Db 799 VAVEAKGGLCYLHSDSPILHRDVSNMILLSDPEAHVADFGAKFLVDGAASECMS 858  
 QY 1046 LAGTPGVPEYQSPFCSTKGDVSYGVLLLELTGKRPTSPDFGDN-NLVGWYKQHA 1104  
 Db 859 IADSYGIAPEYAVYTLKVDKSDVSGVLLLELTGKRPTSPDFGDN-NLVGWYKQHA 915  
 QY 1105 KAIRSDVPELMEKE--DPALE---IELLOHLKVAVACLDRAKRPMTVOYMA 1154  
 Db 916 EEBITQSDAIVAVIADPRLTGYPILTSVHVFKIAMCVBEBAARPTMREVVM 971

RESULT 5  
 US-09-228-986-72  
 / Sequence 72, Application US/09228986  
 / Patent No. 6359198  
 / GENERAL INFORMATION:  
 / APPLICANT: Strabala, Timothy  
 / APPLICANT: Nieuwenhuizen, Niels  
 / TITLE OF INVENTION: Compositions isolated from Plant Cells  
 / FILE REFERENCE: 11000/1020  
 / CURRENT APPLICATION NUMBER: US/09/228, 986  
 / NUMBER OF SEQ ID NOS: 130  
 / SOFTWARE: FastSeq for Windows Version 3.0  
 / SEQ ID NO 72  
 / LENGTH: 910  
 / TYPE: PRT  
 / ORGANISM: Pinus radiata  
 / -228-986-72

Query Match 18.1%; Score 1119.5; DB 4; Length 910;  
 Best Local Similarity 30.5%; Pred. No. 3.1e-73;  
 Matches 332; Conservative 166; Mismatches 365; Indels 227; Gaps 32;

QY 9 LSVTTLPFFSF---FSLSPQAS-----PSQSLREIHQISTFDVL---PDKNLLPDMS 56  
 DB 2 LLLATLSFILFLNPFAPSTPIAHPFHPHPRITNADQHALAFKSAITVDPQSILATSWL 61  
 QY 57 SNNKPCTFDGVC--RDDKVTSIDLSKPLNVGFSVASSLSLTGLESILFINSNINGS 114  
 DB 62 PNVSFCCMTGIGICRRRQRYISLNVSMGLQ---GTTSPILANLSLTYVDLHN----- 112  
 QY 115 VSGFKCA-----SLTSLDSRNSLSGPTVTTLSGSCSGKFLNVSNTLDPFGKY 166  
 DB 113 -NSFDGIRPVQGLTFLKMLRLSKNQLQGISIP--PTLANCSRLRLTISFNNL--TGN 167  
 QY 167 SGGI-KLNSLEVLDSLANSISGANVGVVSDCGSLKHLAISGNKISDVVSCVNE 225  
 DB 168 PPOCLILPNLIGSLGNNITG-----TTPDLGNIS-----SLQ 202  
 QY 226 FLVSSNNSTGIPF-LGDCSALQHLDISGNKLSGDFRAISTCTELKLLNITSNQFVP 284  
 DB 203 YLSLSQNLQGSVSELSGLIIVDLFGNHLTGCIPESSLNCTNBLIDGNQVGH 262  
 QY 285 IP-----PLPKSLQVLSLAKNKFTGSIIPFLGACDTLTGLDLSGNHFGAVPPFGSCS 340  
 DB 263 IPSHLCTKKTQMLMYRLGANNQSGSVPSLFPN-CTKLOEIALPNQLSGIYPMELGKLT 321  
 QY 341 LLESIALSSNNF-SG---ELPMDTLL-KMRGLKVLDSLNEPSGSLPESLNTLSASLITL 395  
 DB 322 HLOQLPFGGNYFISGWTMCPILILNLSGSDIQLYDLSNNFTGQLPSISHLSTRYHL 381  
 QY 396 DLSNNFSPILPNLCQNKNTLQELIYLNNGFTKPIPTLSCSELVSLHSPVYLSGT 455  
 DB 382 DLSNNELA-----GEIPPAIGNSSLFFANLGNNTYGS 415  
 QY 456 IPSSLGSLSKLDLKLMLNLEGEIPORIMVYKLTETLIDPNULTGIPBGLSNTNIN 515  
 DB 416 IPSSLIMQKLERIVMDSNNLQGNIPMEIGQKSLIGLILYLSGNNLSGKIPDVANLQOLR 475  
 QY 516 WLSLNNRLTGEIPKWIQLENLALIKLSNNSFSGNIPDELGDRSL-IWLDLNTNLFNG 574  
 DB 476 YLILNHNQLSGDIANNLGCVALLLDLSYNNLSCGHIQELAGLANLAFYNNLSNNLSG 535  
 QY 575 TTPAMFKQSGKIANFLAGKRYIVYIKDKKKECHAGNLLFEQGISRSQNLNLSRNP 634  
 DB 536 HYPLELGR----- 543  
 QY 635 CHTSRVYTGHTSPFDNNGSNMFLDMSYNNLSGYIPEKISGMPYLFILNIGHNDISGSI 634  
 DB 544 -----FD---MLQAIIDISANGITGYIISYISGMEKAVAYLMLSTNLEGP 585

QY 695 PDEVDLNGILNILLSSNKLDGRIPQMSALTMLTEIDLNNNLGPIPEMQOFETPPA 754  
 DB 586 PVSISELSLQDLDLSSNNLSGCIPISLANLTMHLNFSFNLDSBPVEGFKXIGAT 645  
 QY 755 KFLNPPGLCG--YPLPCDDPBNMGVYAHQRS--HGRRPASLAGSVAMGLFSEVCIFGL 810  
 DB 646 AFMGNLGICGFWNLPPC-----YAHKRSVNLNKRVIILVVVAIVL-----CLFLA 694  
 QY 811 ILVGENMKR-RRKKEALEMYAECHGNSGDPRTANNNTWLTGVKALGINLAAFEKPR 869  
 DB 695 ILMKNCERNIQRDIGPSLNV--GH-----R 718  
 QY 870 KLTFRADLQATNGFHANDSLGSGFGEDVYKAILKDSAAVAILKLHVSGQ-DREFMAEM 928  
 DB 719 RISTAEVLATNEFSDANLIGISFGKVKYKILINDGTWAV-KLNLQNEGAQKSPDRRC 777  
 QY 929 ETIKIKRMLVPLGVCYKVDERLLVNEVMKYSLEVDLQDPKKGAVLXSTRKIAI 988  
 DB 778 KVLGRVHRRLIVITTCYSDLOIKALIFPLMRKSLDKWLYPDDEGQSLNLIORLNTAI 837  
 QY 989 GSARGLAFLHNSCPHIIHRDMKSSNVLDENLEAVSDFGMARLMSAMDT--LSYST 1045  
 DB 838 DIAQMTYLLHHGFQVYIHCDLKPNVLLGEMWTAVLIDFGIATICFANNEDGALTSTNA 897  
 QY 1046 IAGTPGVPP 1055  
 DB 898 LKSGTGYIP 907

RESULT 6  
 US-08-567-375-4  
 / Sequence 4, Application US/08567375  
 / Patent No. 5952485  
 / GENERAL INFORMATION:  
 / APPLICANT: Ronald, Pamela C.  
 / APPLICANT: Wang, Guo-Liang  
 / APPLICANT: Song, Wen-Yuang  
 / APPLICANT: Szabo, Veronique  
 / TITLE OF INVENTION: Procedures and Materials for Confering  
 / NUMBER OF SEQUENCES: 16  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Townsend and Townsend and Crew LLP  
 / STREET: Two Embarcadero Center, Eighth Floor  
 / CITY: San Francisco  
 / STATE: California  
 / COUNTRY: USA  
 / ZIP: 94111-3834  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patentin Release #1.0, Version #1.30  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/567,375  
 / FILING DATE: 04-DEC-1995  
 / CLASSIFICATION: 800  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 60/004,645  
 / FILING DATE: 29-SEP-1995  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/475,891  
 / FILING DATE: 07-JUN-1995  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/373,375  
 / FILING DATE: 17-JAN-1995  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Baectian, Kevin L.  
 / REGISTRATION NUMBER: 34,774  
 / REFERENCE/DOCKET NUMBER: 023070-058930  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1025 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-567-375-4

Query Match 17.6%; Score 1087; DB 2; Length 1025;  
 Best Local Similarity 30.8%; Pred. No. 8.6e-71;  
 Matches 323; Conservative 140; Mismatches 371; Indels 216; Gaps 28;

QY 231 SNNFSGI--PFLGDSALQHDISGNKLSGDSFRAISTCTELKLNINSSNOFVPIPL 288  
 DB 88 SSNLS-GIISPSLGNISFREDLGNVLSGELPELSLSKOLLESDNSIGSIFA- 145  
 QY 289 PLKSLQYLIAENKFGELPFLSGACDTLTGLDSGNHFGAVPFPQ-SCSLIESIAL 347  
 DB 146 -----AIGACTKLTSLDSHNLQRMIPREIGASIKHLSNLYL 183  
 QY 348 SSNNFSGELPMDTLKKRGLKVIDLSPNPSGELPESLTNLSASILTLDLSSNNFSGPIL 407  
 DB 184 YKNGLSGEIP-SALGNLTSIQEPLDLPNRLSGAIPSLQSL-SLTNNVLGQNNLSG-MI 240  
 QY 408 PNLQCN-----PKNTLQELYL-----QNNGFYKLIPTLSNCELY 443  
 DB 241 PMSIWNLSLRAFSVRNKGMLPTNAFKTLLILEVIMGNTRFPGKLPASVANAHLT 300  
 QY 444 SLHLSFNYLSGTPSLSGSLSKDLKLMNM----- 475  
 DB 301 VIQYGNLFSGITTSFGRLEMLTELAMNLPTQREDDWGFISDLNCSKLTQTLNCE 360  
 QY 476 --LGEIQLMAYKT-LETLIDPNDLGEIPSGLSNCTNLMISLNNRLTGEIPKI 532  
 DB 361 NNLGVLNPSFSNLSSTLSFALBELKNTGSIPIKDIGNLGLQHYICNNFPGSL 420  
 QY 533 GLEMLATLKLSNNSFSGNIPDELGDCKSLIMDLNLTNLTNGTTPAAMEKQSG---KIA 588  
 DB 421 GRKMLGILLAVENNLSSISPLAIGNTELINILGTNKSQWIPYLSMTMLSLIGS 480  
 QY 589 ANFIAGKRYVYIKDGMKKECHGAGNLEFGISSEQLNLTNPENITSRVYGHTSP 648  
 DB 481 TNNLSGP-----IPSELFNIGTSLIMNVSKNLSSIFQ 515  
 QY 649 TEDNNGSMMFLMSYNMISGYIPKEISGMPYLITLNGHNDISSIPDEVDLREGNLTLD 708  
 DB 516 EIGHLKNLVEFHAESNRLSKIPNTLGDQQLIRYVYLNNNLSSISALGQLETLTD 575  
 QY 709 LSSNKLGRIPQAMGALTMLTEIDLSNNNLGPIPEMGQFETPPAKFLNPGICG---- 764  
 DB 576 LSSNNLSGQIPSLADITMLHSLNLSFNSFVEGVEPTTAFMAAGISIGQAKCGSIPD 635  
 QY 765 YLPREDPSNAGQYAHQSHRBPASLAGSVAMGLFSPFCIFGLILVGRMKRRRK 824  
 DB 636 LHLPRCCP-----LLENKHFPEVLPISVSLAALAILSSVILI-----TWKRTKK 682  
 QY 825 EAELEMYAEGHNGSDRTANNNTWKLTVYKALSLINLAFFKPLKLTPLADLLQNTGPH 884  
 DB 683 GAPSRTSMGH-----PL--VSQSOLVYATGFA 709  
 QY 885 NDSLISGGGFGDYKAILKDGSAVAIKKLIHVSOGDREPMAMETIKIKHRLNVPPLG 944  
 DB 710 PTMLTSGSGFSGYKGLINIDHVAVKVLENPALKSFTEBCALNNMRRLVKIVT 769  
 QY 945 YCVKGDG-----LVNEVWKYSLDVLQ---DPKKGVKLKLSTRKIALGSRGLAF 996  
 DB 770 TCSIDNRKNDFAKIYIDFNPNSLEDWIMHPTNQADGRHLNRRVYITLLDVACALDY 829  
 QY 997 LHNHCSPHIIRDMKSSNVLLDENLBARVSDFGMARLM---SANDTHLSVTTLAGTGY 1052  
 DB 830 LHRHGPFFVHCDIKSSNVLLSDDMVAHVGDPLAIRIIVDGTSLIQOSTSSMGFITGTY 889

QY 1053 VPEYQSPFCSTGKDYVYGVVLELTKGKPTDS---PDFGDNVLGVWQKHAQL--- 1106  
 DB 890 AAPYGVGLASTHGDYVGLVLETVGKPTDSTPRDLG-----LQYVELGLH 942  
 QY 1107 -RISVDPPELM-----KEDPALRE--LLOHLKVAACLDL-RAMRPYMWQVM 1152  
 DB 943 GRVTVVDTKLIDSENMWLNSTNNSPCRRITECTIVMLRLRGLSCQELPSSRPT----- 997  
 QY 1153 AMFEKIQAGSIDQSTIRSTEDGGFSTIE 1182  
 DB 998 -----GDIIIDELMIAIKONTLSGLFVCE 1019

RESULT 7  
 US-08-587-680A-4  
 Sequence 4, Application US/08587680A  
 Patent No. 5977434  
 GENERAL INFORMATION:  
 APPLICANT: Ronald, Pamela C.  
 APPLICANT: Wang, Guo-Liang  
 APPLICANT: Song, Wen-Yuang  
 APPLICANT: Sabo, Veronique  
 TITLE OF INVENTION: Procedures and Materials for Confering  
 TITLE OF INVENTION: Disease Resistance in Plants  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/587,680A  
 FILING DATE: 17-JAN-1996  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/373,375  
 FILING DATE: 17-JAN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/475,891  
 FILING DATE: 07-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/004,645  
 FILING DATE: 29-SEP-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/567,375  
 FILING DATE: 04-DEC-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Baebian, Kevin L.  
 REGISTRATION NUMBER: 34,774  
 REFERENCE/DOCKET NUMBER: 023070-058940US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1025 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-587-680A-4  
 Query Match 17.6%; Score 1087; DB 2; Length 1025;  
 Best Local Similarity 30.8%; Pred. No. 8.6e-71;  
 Matches 323; Conservative 140; Mismatches 371; Indels 216; Gaps 28;  
 QY 231 SNNFSGI--PFLGDSALQHDISGNKLSGDSFRAISTCTELKLNINSSNOFVPIPL 288



```

Db      88  SSNS-GTISPSLGNLSFLRLDIGNVLSGELPPELSRLQLLELDNSIQGISPA- 145
Qy      289 PLKSLQVLSLAENKFTGEIPDPFLSGACDTLTGLDLSGNHFYAVPPFG-SCSLESIAL 347
Db      146  -----AIGACTKLTSLDLSHNLQRLGMIPEIGASLKHLSNLT 183
Qy      348 SSNNFSGELPMDTLTKRKGLKVLDSFNEFSGELPSLTNLSASLLTDLSSNNFSGPL 407
Db      184 YKGLSGEIP-SALGNLTSIQEFDLSFNRLSGAIPSSLQOLS-SLTTNMLGQNNLSG-MI 240
Qy      408 PNLCON-----PKNTLOEYL-----QNNGFTGKIPTLSNCSLEY 443
Db      241 PMSIWNLSLARAVRENKLGMIPTNAFKTLHLLEVIMDGNTRFGKIPASVANASHLT 300
Qy      444 SLHSFYVLSGTIPSSLSLSKLRDLKMLM----- 475
Db      301 VIQYGNLFGIITSGFGRRLNLTLYLWRNLQTRBQDDWGFISDLTNCSLQTLNIGE 360
Qy      476 --LEGELPQELMYKT-LETLILDPRDLGELPSSLSNCTNLSNNRLTGEIPKI 532
Db      361 NMGIVLPSFSLSTLSFLALEINKITGSIPIKIGNLIGLOHLYLCNNFRGSLPSSL 420
Qy      533 GRLENTALTLKSNNSFSGNIPDELGDCRSILWLDNTNLNFGTIPAAFKQSG-----KIA 588
Db      421 GRLENTALTLKSNNSFSGNIPDELGDCRSILWLDNTNLNFGTIPAAFKQSG-----KIA 588
Qy      589 ANFIAGKRYVYIKNDQMKCEGAGNLEFGQIRSEQNLRLSTRNPNITSRYVGHSTP 648
Db      481 TNNLSGP-----IPSELFNIGLTSIMINVSNNLEGSIPQ 515
Qy      649 TFDNNSGMFLDMSYMLSGYIPKEIGSMPLYFLINLGNDSISIPEDVDLGNLTL 708
Db      516 EIGHLKNVLPFAHESRSLGKIPNTLIGDCQLRYLYLQNNLSGSIPLAQOLGLETLD 575
Qy      709 LSSNKLGRIPQAMSAITLWLEIDLSNNNLSGPIPEMGOFTFPFAKFLANPGCG----- 764
Db      576 LSSNKLGRIPQAMSAITLWLEIDLSNNNLSGPIPEMGOFTFPFAKFLANPGCG----- 764
Qy      765 YLPREDPENNAGYVAHQSHRRPASLAGSVAMGLFSPVCIFGLILVGREMKRRRK 824
Db      636 LHPRCPCP-----LLENRRGFVLPISVSLAALAILSLYLLI-----TWKRTKK 682
Qy      825 EAELEMYAEGHNSGDRANTNTMKTLYKALSLINLAPEKPLKLTFRADLLQATNGFH 884
Db      683 GAPRTSMKGH-----PL--VSYSQLYKATDGF 709
Qy      885 NDLISGGGFGDVYKAILKDGSAVAIKKLHVSGQGRFPAEMETIGIKHRLVPLLG 944
Db      710 PYNLLSGSFGSVYKGLNIPDHVAVKYLKLENPKALKSFAECEALRNRRRLVIVT 769
Qy      945 YKVGDER-----LLVNEVMKYGSIEDVLO--DPKKGCVKLTSTRKIALSGARGLAF 966
Db      770 ICSSISDNRGNPKALVYPMFNGSLEDWIMHETNDQADQRHLNHRVTLILLDVAACLDY 829
Qy      997 LHHNCSPIIHRDMKSSNVLLDENLEAVSDFGMARLL-----SAMDTHLSVSTLAGRGV 1052
Db      830 LHRHGEPEPVVHCDIKSNVLLDSDMVAVHGFGLARILVDGTSILQOSTSMGFTIGTIG 889
Qy      1053 VPEPYQFSGSTKDVYSYGVVLELLTGRPTDS--PDFGDNMLVGVWKQAHKL-- 1106
Db      890 AAPFYGVLIASTHGDIVSYGILVLEIVTGRKRPDTSFRRPDLG-----LRQYVELGHA 942
Qy      1107 -RISDPPELM-----KEDPALEIE--LLQHLKAVACLD--RAMRPPYVQW 1152
Db      943 GRVYDVDTKILDSKWNSTNNSPCRRITECTIVMLRLGLSCOSLPSSRPT----- 997
Qy      1153 AMFKEIQAGSIDQSTIRSIEDGFSTIE 1182
Db      998 -----GDIIDELNMAIKONLSLFPVCE 1019

```

RESULT 8

```

US-08-475-891A-4
; Sequence 4, Application US/08475891A
; Patent No. 5859339
; GENERAL INFORMATION:
; APPLICANT: Ronald, Pamela C.
; APPLICANT: Wang, Guo-Liang
; APPLICANT: Song, Wen-Yang
; TITLE OF INVENTION: Procedures and Materials for Confering
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,891A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/373,375
; FILING DATE: 17-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baselian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 02370-058910US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1012 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1012
; OTHER INFORMATION: /note="Xa21 Xanthomonas spp.
; OTHER INFORMATION: disease resistance polypeptide RRK-B
; OTHER INFORMATION: from rice (Oryza sativa)"
US-08-475-891A-4
Query Match 17.4%; Score 1073.5; DB 2; Length 1012;
Best Local Similarity 30.5%; Pred. No. 8.2e-70;
Matches 320; Conservative 139; Mismatches 362; Indels 229; Gaps 29;
Qy      231 SSNNFSGI--PFLDSCALQHLDISGNKLSGDFSRASITCTELKILNLSNNGFVGPPL 288
Db      88  SSNS-GTISPSLGNLSFLRLDIGNVLSGELPPELSRLQLLELDNSIQGISPA- 145
Qy      289 PLKSLQVLSLAENKFTGEIPDPFLSGACDTLTGLDLSGNHFYAVPPFG-SCSLESIAL 347
Db      146  -----AIGACTKLTSLDLSHNLQRLGMIPEIGASLKHLSNLT 183
Qy      348 SSNNFSGELPMDTLTKRKGLKVLDSFNEFSGELPSLTNLSASLLTDLSSNNFSGPL 407
Db      184 YKGLSGEIP-SALGNLTSIQEFDLSFNRLSGAIPSSLQOLS-SLTTNMLGQNNLSG-MI 240
Qy      408 PNLCON-----PKNTLOEYL-----QNNGFTGKIPTLSNCSLEY 443
Db      241 PMSIWNLSLARAVRENKLGMIPTNAFKTLHLLEVIMDGNTRFGKIPASVANASHLT 300
Qy      444 SLHSFYVLSGTIPSSLSLSKLRDLKMLM----- 475

```

Db 301 VIQYGNLPSGDIITSGRGLRNLTELYLWNLFTQREODDNGFISDITNCSKLOTLNIGE 360  
 Qy 476 --LEGEPOSLMAYKT--LETLILDPNDLGEIPSGLSNCTNLMWISLNNRLTGEIPKMI 532  
 Db 361 NNNGVLPNSFSLSTLSFLALELNLKTSIKDIGNLGLHVLICNNRFGSLPSL 420  
 Qy 533 GRLENTLALIKLSNNSFSGNIPDELGDCRSILWLDLNTLNGTIPAMFKQSG---KIA 588  
 Db 421 GRKLKGLILAYENNLSSGLPLAIGNLTELINLLGTNKSFGWIPYTLISLTLNLGLS 480  
 Qy 589 ANFLAGRRYYIKNDGKKCKGAGNLEFQGRSEDLNLSLRNCPNITSRYGGHTSP 648  
 Db 481 TNNLSGP-----IPSELFNTQTLSTIMINVSKNLBSITQ 515  
 Qy 649 TFDNNGSMFLDMSYNNLSGYIPKEISMPYLFIPLNGNDISGIPDEVDLRLNILD 708  
 Db 516 EIGHKLVLFHNASNLSSKIPNTLGDCCOLRLYIQLNLLSGSIPALGQKGLETLD 575  
 Qy 709 LSNKLDGRIPQAMSAITMLTEIDLNNNLSGPIEMGQETTPPAKFLNPGLCG---- 764  
 Db 576 LSSNNLSGQIPSTLADITMLSLNLSFNSFVGEVPTIGAPAAAGSISIQNAKLCGIPD 635  
 Qy 765 YPLRCDSNADGYAHQORSHGRPASLAGSVAMGLFSPVCIPLGLILYGRERKKRKK 824  
 Db 636 LHLPRCCP-----LLENKHPVLPISVSLAALAILSLYL-----TMKRTTK 682  
 Qy 825 EAELEMYAEGHNGSDRTANNTNWKLTGVKEALSINLAPEKPLKLTFLADLQATNGFH 884  
 Db 683 GAPSTSKNGH-----PL--VYSQLVKATGFA 709  
 Qy 885 NDSLIGSGFDVYKAILKDSAVAIKLIHVSQGDREFAEMETTKIKHRLVPLLG 944  
 Db 710 PTNLIGSGFSFVSVKKLENPAL-----KSFIAEEALRNHRLVPLVT 756  
 Qy 945 YCKVDER-----LLVNEVMYKGSLEVDLQ---DPKKGVLKLTSTRKIAIGSAGIAF 996  
 Db 757 ICSSIDNKGNDPKAIYDPMFNGSLJEDWIHPETNOADQRLNHRRTYLLDVAICALDY 816  
 Qy 997 LHHNCSPHIHRDMKSNVLLDENLEARVSDFGMARL---SAMDTLSVSTLAGTGY 1052  
 Db 817 LHRHGPBPVHCDIKSSNVLLDSMAHVGFLARILVDTGSLIQSTSSMGFLTIGY 876  
 Qy 1053 VPPEYQSRCTKGDVYSYGVVLELLTGKRPDS---PDEGDNLVGMVYKONAKL--- 1106  
 Db 877 AAPEGVGLIAGTHGDIYSYGLIVEYTKKPTDSTRPDLG-----LRQYVELGLH 929  
 Qy 1107 RISVDFPELMA-----KEDPALEIE--LLQHLKVAVALCD--RAMRRPTMVQM 1152  
 Db 930 GRVTVVDVTKLILDSNNMLNSTNNSPCRTITBCIYWLTLGLSCQELPSSRTP----- 984  
 Qy 1153 AMFEIQAGSGIISOSTIRSIEDGFSSTIE 1182  
 Db 985 -----GDIIDELNAIKONLSGLFPVCE 1006  
 RESULT 9  
 US-09-228-986-73  
 ; Sequence 73, Application US/09228986  
 ; Patent No. 6359198  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Strabala, Timothy  
 ; APPLICANT: Nieuwenhuizen, Niels  
 ; TITLE OF INVENTION: Compositions Isolated from Plant Cells  
 ; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling  
 ; FILE REFERENCE: 11000/1020  
 ; CURRENT APPLICATION NUMBER: US/09/228,986  
 ; CURRENT FILING DATE: 1999-01-12  
 ; NUMBER OF SEQ ID NOS: 130  
 ; SOFTWARE: FastsEQ for Windows Version 3.0  
 ; SEQ ID NO 73  
 ; LENGTH: 947  
 ; TYPE: PRF  
 ; ORGANISM: Pinus radiata

US-09-228-986-73  
 Query Match 17.1%; Score 1057; DB 4; Length 947;  
 Best Local Similarity 30.5%; Pred. No. 1,2e-68;  
 Matches 331; Conservative 164; Mismatches 377; Indels 214; Gaps 36;  
 Qy 137 PVTTLTSLGSC--SGLELVNVSNTLDPFGKVGGLKLSNLEVLDSANSIGANVGVY 194  
 Db 9 PFLSLPLTSLDSBGTCLDQOQYVWK-----KLSYTP---SWTVKSDNCGWS 56  
 Qy 195 LSDGCGE---LKHIAISGNKISGDVDSRCVNLFLDVSSNNSTGTPIFGDCSALQHL 251  
 Db 57 GVNCTAESNVTEHLMGSGFRKGNAMQITC-----KIQALQVLD 95  
 Qy 252 ISGNLSCDPFRAISTCTELKLNISNNOFYG--PIPL-PLKSHQVLSLENKFTGEIPD 309  
 Db 96 VSDNLSTPSDNDIOACTNPLFLNISNPLPGSSLPLAEMRKHLFDVHNGAFGEFGP 155  
 Qy 310 FLGACDTLTGCLDLSGNHFYAVPPFGSCSLLESLSLSSNNSFGEILPMOTLKMRLKV 369  
 Db 156 QIQHLLTD--LRVLNLTNNFSGPIPSFLGNLTTLTEKIDFSGQYFGEPEPKK--LVACTSLTY 213  
 Qy 370 LDLSFNERSGELPESLNTLSALTLTDLSSNNSGPI-----LPINQMPKXTLQELYL 423  
 Db 214 LDLSFNRRLTGQIPDNI SNL--IHEFLILSSNNLTGTPKTLDRLVNLTFRASN----- 265  
 Qy 424 QNNGFTGKIPPTLSNCSLVSLSHLSFNYSGLTIPSSLSLSKLDLKLMLMGEIPQ-- 482  
 Db 266 -KQQLIGRIPVOLAKLTFLHFLDSLNGNLNETIPELPALSNLQTLDTKRLTGEIIPQN 324  
 Qy 483 ---ELMYVKTLETLI-----LDEND--LGEIPSGLSNCTNLMWISLSN 521  
 Db 325 FSRKILIRLRIGQNLKGNIPLTIGNMSNLTYLEMNDNSLDQIPLQALNCKILDLIDGN 384  
 Qy 522 NKLTEGIPKTRLENLALIKLSNNSFSGNIPDELGDCRSILWLDLNTLNGTIPAMF 581  
 Db 385 NNLSSGLTNQLPSLQLOVYTKLHNNFVGSIPYILSSSNLSYDLSNTLNGSIPSNIF 444  
 Qy 582 KQSGKIAANFIAGKRYVYIKNDGKKCKGAGNLEFQGRSEDLNLSLRNCPNITSRV 641  
 Db 445 -----NLSKQLNLR--QNNKLTGAILPNV----- 467  
 Qy 642 YGHTSPPTDNNGSMFLDMST--NMLSGYIPKEISMPYLEI--LNLGNDISGIPDEV 698  
 Db 468 -----GGSQVLELLOLGNNLTGTMPLTEIGPARKLOIQLNLSNLSGEGIPNTL 516  
 Qy 699 GDLRGNIILDSNKKLDGRIPQAMSAITMLTEIDLNNNLSGPIEMGQETTPPAKF-- 756  
 Db 517 SGLYMLIILDSNKKLTGEVPSGLTLMLSLTILNLSNLSLTGLPK-----FPMSTSL 570  
 Qy 757 ---LNNPGLCGYPLPRCDPSNADGYAHHQRSHGRPPASLA-----GSVAMGLLF 802  
 Db 571 IILDGNPGLTA-----GQNSAPASASARKKISAILIIGVAVAGAVF 612  
 Qy 803 SFVCIIGLILYGRM--RKRKKEALEMYAEGHNSGSDRTANNTNWKLTGVKEALSIN 860  
 Db 613 AIVAV--GLFTVASKYFGREDQMPVOLARKTBGHTHPDSI-----HRLRID 659  
 Qy 861 LAPEKPLKLTFLADLLOATNGFHNDSLIGSGGFGDVYKAILKDSGAVAIKKL-----IH 915  
 Db 660 ---FEKGVFA-----TLDPANVFLKXK-----FSTYYKAVMPSGISYGVKLMWSDRIF 705  
 Qy 916 VSGGDREFAEMETTKIKHRLVPLLGCKYKVDRLVNEVMYKGSLEVDLQDPKKG 975  
 Db 706 KSG--SYRKLGHELEKQKLRHPNLPPLAHVLDTPDSAYFEYEVHNGSISEFHTSNVS- 763  
 Qy 976 VKKLSTRKIAIGSAGLAFLHHNCSPHIHRDMKSNVLLDENLEARVSDFGMARLMS 1035  
 Db 764 -VLDMPRCRILAGVAGGLAFL--HGCHPIPHLDITTKAILLSIRPQIGDELBCKIYD 821  
 Qy 1036 AMDTHLSVTLAGTGGVVPPEYQSFRCSTKGDVYSYGVVLELLTGKRPDSDPFGDN 1095  
 Db 822 PSKSTGISALAGSVGVPPBYAVTMKVTAAANVYSFVILTEILGRTPTS-----GMD 877

QY 1096 LVGWYKQ--HAKRISDVDPDELMKEDPALEIELEIOLKVAACIDDDRAMRPTWQYMA 1153  
 Db 878 LAKWOSTLSGEEWEQIDTGINFNSVOIQNEMTAMKVALSCVSSSPESRPMKRNWVG 937  
 QY 1154 MEKEIO 1159  
 Db 938 MLOWMR 943

## RESULT 10

US-09-353-585-2  
 ; Sequence 2, Application US/09353585  
 ; Patent No. 6287865

## GENERAL INFORMATION:

APPLICANT: Dixon, Mark S  
 Jones, David A

TITLE OF INVENTION: Plant pathogen resistance genes and uses  
 thereof

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon & Vanderhye PC  
 STREET: 8th Floor, 1100 No. 6287865th Glebe Road  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: United States of America  
 ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/353,585

FILING DATE: 15-Jul-1999

CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q

1/68

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/930,277

FILING DATE: 27-OCT-1997

APPLICATION NUMBER: PCT/GB96/00785

FILING DATE: 01-APR-1996

APPLICATION NUMBER: GB 9506658.5

FILING DATE: 31-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Ms Mary J Wilson

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 620-69

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1112 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

MOLECULE TYPE: linear

HYPOTHEICAL: YES

ORIGINAL SOURCE:

ORGANISM: Tomato

STRAIN: CF2

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-353-585-2

Query Match

Beet Local Similarity 14.4%; Score 891; DB 4; Length 1112;

Matches 283; Conservative 129; Mismatches 329; Indels 184; Gaps 30;

QY 74 VTSIILSPFLNWSAVSSLSLTGLESFLNSHNGSVSGPK-SASITSIDLSRN 132

Db 217 LTELDSIDMAN---GSFASLIGNMNNISFLPLNGNQLSGSIPETCYLRSLTYLDISEN 273

QY 133 SLGSPVTLTSLGSGSGELKELNVSNTLDFPGKVGGLK-----LNSLEVDLSANSLG 187  
 Db 274 ALNGSLP--ASLGNLNNLSFLFLYGNQ-----SGSLPEEIGYLRSLNVLGLSENALNG 325  
 QY 188 ANVVGWVLSDGGELKHLA---ISGNKISGDVDS--RCVNEFLDVSNNFSTGIP-FL 241  
 Db 326 S-----IPASIGNLKNLSRLNVLNNGSLGSLPALSIGNLNNLSMLTYNNQISGSLP 379  
 QY 242 GDCSALQHDIDSGKLSGDFPSAISTCTELKLNLSNQFVGPVPL-PLKSLQYLSLA 299  
 Db 380 GNINLNSMLTYNNQISGSLPALSIGNLNNLSRLTYNNQISGSLPEEIGYLSLTYLDS 439  
 QY 300 ENKFTGEIPD-----FL-----SGACDTLTGLDLSGNHFGVAVPEPF 336  
 Db 440 NNSINGFIPTASFGMNSNAFLFLYENQLASSVPEITGLRSLNVLNVDLSNNALNGSLP 499  
 QY 337 GSCSLIESLALSSNNFSGELPMDTLNKRGLKVLDSFNEFSGELPESLTJMSASLTLD 396  
 Db 500 GNINLNSRLNVLNNGSLIP--EEIGYLRSLNVLNVDLSNNALNGSLPASFNLNLSRLN 557  
 QY 397 LSSNPSGPILENLCONKNTLQELYLONNGFTGKIPTLSNCELYSLHSFVYLSGTI 456  
 Db 558 LVNNQISGSLPEE--GYLRSLNVDLSNNALNGSLPALSIGNLNNLSMLTYNNQISGSI 615  
 QY 457 PSLGSLKRLDLKMLNMLEGIPQELMYKTLTETLIDFNDLTGEPISGLSNCTNLMW 516  
 Db 616 PEIGYLSLTYLSSGNNSLNGLIPASFRANKVLOALINNNLIGELPSSVCNLTSLV 675  
 QY 517 ISLSNNRLTGEIPKWTIGLENLAILKUNNSFSGNIDDELDCSLIWLDTNLTLPNGTI 576  
 Db 676 LYPENNLKGVYPCQCLGNISMLQYLSMSNSFSGELPSSISNLTLQIDGRNNLEGAI 735  
 QY 577 PAAM-----FKSGKTAANPFIAGKRYVYIKNDGKKE-----CH----- 610  
 Db 736 PQCFNNSLSLEVPQNNKSLGTLPTNPSIGCSLISLNLGNELEDBIPSLDNCKLQV 795  
 QY 611 -----GAGNLEFGQIR--SEOL-----NRLST 631  
 Db 796 LDLDGNQNDTFFPMWGLTLPBLRVLRLTNKLNHPIRSSRAELMFPDLRTITDSRAFSQ 855  
 QY 632 RNPCLNTRVYVGGHT-----SPFTDN--NGSMW-----FLDSYNN 665  
 Db 856 DLPTSLFEHLKQRTVIXTMBEPEBSYDSDVYVYVTKGLEIVRLISLYTVYIDLSNK 915  
 QY 666 LSGYIPKEIGSMPLFIPLNGLNDISGIPDEVDLRLGNIILDSKDKLRIPOMSAL 725  
 Db 916 FECHTPEVLDLIRILNVSNNALQYIPSSLSLSTLSDLSFNQISGELPQOLASL 975  
 QY 726 TMLTEIDLSNNLSGPIPEMGOETFPFAFLNPNGLCGYPLPR-C--DPSNADGYA--- 779  
 Db 976 TFLFELNLSHNYLQGCIPQCPQRTFESNGBEGNDGLRGVFGKGGKXDVSKRYTVA 1035  
 QY 780 -HHQRSHGR-----RPASLAGSVAMGLPFCVIGILV-----REMRK 820  
 Db 1036 LBDDESNSEFFNDFWKALWG--YSGGCGIGISMTYILISGNLRLARLIEKLEHRTIQ 1094  
 QY 821 RKKKAELEMYAGHGSGSDRTANN 845  
 Db 1095 RKKQ-----RQGRNRYRRNN 1110

## RESULT 11

US-09-180-439-8

; Sequence 8, Application US/09180439  
 ; Patent No. 6225532

## GENERAL INFORMATION:

APPLICANT: Dixon, Mark S  
 Jones, David A

APPLICANT: Hatzixanthis, Kostas  
 APPLICANT: Jones, Jonathan DG

TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof  
 ; FILE REFERENCE: 620 - 53

CURRENT APPLICATION NUMBER: US/09/180,439  
 EARLIER FILING DATE: 1998-12-06  
 EARLIER APPLICATION NUMBER: PCT/GB97/01249  
 EARLIER FILING DATE: 1997-05-08  
 EARLIER APPLICATION NUMBER: GB 9609661.3  
 EARLIER FILING DATE: 1996-05-09  
 EARLIER APPLICATION NUMBER: GB 9619924.5  
 EARLIER FILING DATE: 1996-09-24  
 NUMBER OF SEQ ID NOS: 10  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 8  
 LENGTH: 1016  
 TYPE: PRF  
 ORGANISM: Lycopersicon esculentum  
 US-09-180-439-8

Query Match 14.4%; Score 890; DB 4; Length 1016;  
 Best Local Similarity 29.0%; Pred. No. 2e-56;  
 Matches 304; Conservative 146; Mismatches 352; Indels 248; Gaps 38;

2 KTESFFLSTVTLTFPFSSLSFQSPSSQSLYREIHQISFKDVLPRKN-LLEPWSNK 59  
 7 KVFSS--LQFTVY--LFTYAFST-----EATALLMKKATFKQNNSSFLASWTSS 56  
 60 NPC-TFDGVTGRDVKYTSIDLSKPLNVGSVSSLL-----SLGLESLFSSNSHNG 113  
 57 NACDWCWGVVCLNGRVNT-----LNTNASVIGTLVAFPFSSLPFLNLDLSSNNISG 109  
 114 SV-----SGF-----KCSASL 124  
 110 TIPEIGN/TLVYLDLNTNOISGTFIPQISLAKLQIRFNHNLGFLPEBIGYRSL 169  
 125 TSLDSSRLSGPVTTLSLSCSGSLKFLVANSNTLD-----F 162  
 170 TKLSIGLIFLSSIP--ASLGNMTLSLFLYENGLSGFPEBIGYRSLTKSLDINFL 227  
 163 PGKVSGL-KLNSLEVLDSANSISGA--NVGVVLSGCGELKHLAISGKISGDVDS 219  
 228 SGSIPLASIGLNNLSFLYLYNNQLSGSIPEBIGYRSL-----LTKLSIGLIFLSSIPAS 282  
 220 --RCNLEFLDVSSNNFSGTIP-FLGDSALQHDISGNKLSGDFSRASITCTELKINI 276  
 283 LGNINLRLDLYNNKLSGSIPEBIGYRSLTYLDLGENALNGSIPLASIGLNNLPMYL 342  
 277 SSNOFGVPIPLU--PLKSLQVLSLAENKFTGEIPDLGACDTLTGLDLSGHHFYGAVP 334  
 343 YNNQSGSIPEBIGYRSLTYLDLGENALNGSIPLAS--GNNLNLRLDLYNNKLSGSIPE 401  
 335 FPGSCSLLESLSLSSNFGELPMDTLKMRGLKVLDSFNEFSGELPESLTNLSASILT 394  
 402 EIGYRSLTYLDLGENALNGSIPLASIGLNNLPMYLTYNNQLSGSIPEBIGYRSL 459  
 395 LDISNFFSGPI--LPNLQNPKNLTQELLYONNGFTKIPPTLSCSELVSLHSFNY 451  
 460 LHYGNNSLNGSIPLASIGL--NNLPMYLTYNNQLSGSIPEBIGYRSLTYLFLGANS 514  
 452 LSGTIPSSIGSLSKLDLKLMLMTEGELPQBLMYKYLETLLDPNDLTGPIBGLSNC 511  
 515 LNSGIPASIGLNNLSRLYLYNNQLSGSIPLASFGNNRNQTLFLSNDNDLIGIPFVCL 574  
 512 TNLNWSLSGNNRLTGEIPKMGIGLEMLALIKLSNNSFGSNIPDELGDCLSLTMDLNTL 571  
 575 TSLLEVLYMSKNNLKGVPOCLGINSIDHLISMSNSFRGELPESISNLTSLKILDBGRNN 634  
 572 FNGTIP-----AAMF-----KOSGKLAAFLAKRYVYI-----KNDGM 605  
 635 LBGALPOFFGNISLQVFMQNNKSLGTLPNTSICSLISLNLHGNELADEIPSLDNC 694  
 606 KK-----ECHG-----AGNLLPFOGIRSSQANR 628  
 695 KKLQVLDLGNQNDTFPMMVIGTLPELRLVRLTNSNKLHGFIRSSGABLWPPDLRIIDLSR 754  
 629 --LSTRNPNITSRYVGGHT-----SPTFDN--NGSMM-----FLD 660

Db 755 NAFSODLPTSLFELHKMRKIVDKTMEBPSESYVDSVYVVTGKLEIYRILSLTYIID 814  
 QY 661 MSYNMLSGYIPEKIGSNPYFLINLGNHDSGSIPEVEGDLRGINLIDSSNLTDRIQ 720  
 Db 815 LSNKFGHGPSVLGDLIALRVNLVNSHNAQGYIPSSLSGLSILIESLDSFNQLSGIPQ 874  
 QY 721 AMSALTMLELDSNNLSGPIPEMOGFETFPAPKFLNNPGLGYPPLR-C--DPSNADG 777  
 Db 875 QLASLTFLEVNLISHNYLQCTPGQPFRTFESNSYEGNDGLGYPVSKCGCKDPVSEKN 934  
 QY 778 YA-----HHRSHGR-----BPASLGSVAMGLLPSFVCIFGLIVG-----R 815  
 Db 935 YTVSALDEQESNSRFPNDFWKALMG-YGSGLCIGISIIYILISTGNLRLWARIIBELH 993  
 QY 816 EMRRKRRKKEALEMYAEHGNSGDRTANN 845  
 Db 994 KIVQRRKKQ-----RGQRYRRRNN 1014

## RESULT 12

US-09-353-585-3  
 Sequence 3, Application US/09353585  
 Patent No. 6287865

## GENERAL INFORMATION:

APPLICANT: Dixon, Mark S  
 Jones, David A  
 Jones, Jonathan DG

TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:  
 ADDRESS: Nixon & Vanderhye PC  
 STREET: 8th Floor, 1100 No. 6287865th Glebe Road

CITY: Arlington  
 STATE: Virginia

COUNTRY: United States of America  
 ZIP: 22201-4714

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/353,585  
 FILING DATE: 15-Jul-1999  
 CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12O

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/930,277  
 FILING DATE: 27-Oct-1997  
 APPLICATION NUMBER: PCT/GB96/00785  
 FILING DATE: 01-Apr-1996  
 APPLICATION NUMBER: GB 950658.5  
 FILING DATE: 31-Mar-1995

ATTORNEY/AGENT INFORMATION:  
 NAME: Ms Mary J Wilson  
 REGISTRATION NUMBER: 32,955  
 REFERENCE/DOCKET NUMBER: 620-69  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000  
 TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1112 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Tomato  
 STRAIN: Cf2

577 LKILDFGRNNIEGALPQIFGNTSSLOVFDMQNNKLSGLPLPTNFSIGCCSLISLNIHGNEIA 636

QY 601 -----KNDGMRK-----ECHG-----AGNLEF 618  
 Db 637 DEIPRLDNCKKQVLDLGNQNDTFFPMMLGTLPELRVLRILSNLHPISSGAEIMF 696  
 QY 619 QGIRSEQLNR--LSTRNPNITSRYVGGHT-----SPTFDN--NGSNM----- 657  
 Db 697 PDRLITDLSNNAVSQDLPSTLFEHLKGMRTVDKMEEPSEYSDSVVVVTKGLELEIV 756  
 QY 658 -----FLDMSTYMLSGTIPKEIGSMPIYFIINLGHNDISGSIPEVGDRLGNTLIDLS 710  
 Db 757 RLISLTYITIDLSNKPFGHILPSVLGDLAIRVLNVSNNALQGYIPSSLSLSTLSLIDLS 816  
 QY 711 SNKLDGRIPQMSALTMLEIDLSNNLSGPIPEMGQFETFPAPKFLNPGLCGYPLPR- 769  
 Db 817 FNQLSGEIPQOLASLTLEVLNLSHNLYLGGCIPGGPQFRTFESNSYEGNDGLRGYPVSKG 876  
 QY 770 C--DPSNADGYA--HHQSHGR-----RPASLAGSVAMGLFSPFCIFGLIVG----- 814  
 Db 877 CGKDPVSEKNYTVGALDEQESNEFEFNDFWKALMG--YGSGLCIGISIIYILISTGNLRW 935  
 QY 815 -----REMRKRRKKEALEMYAEHGNSGDRTANN 845  
 Db 936 LARIIELEHKIIVORRKKQ-----RQQRNRRRN 966

RESULT 14  
 US-09-180-439-4  
 ; Sequence 4, Application US/09180439  
 ; Patent No. 6225532  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dixon, Mark S  
 ; APPLICANT: Hatzixanthis, Kostas  
 ; APPLICANT: Jones, David A  
 ; APPLICANT: Jones, Jonathan DG  
 ; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof  
 ; FILE REFERENCE: 620 - 53  
 ; CURRENT APPLICATION NUMBER: US/09/180,439  
 ; EARLIER FILING DATE: 1998-12-06  
 ; EARLIER APPLICATION NUMBER: PCT/GB97/01249  
 ; EARLIER FILING DATE: 1997-05-08  
 ; EARLIER APPLICATION NUMBER: GB 9609681.3  
 ; EARLIER FILING DATE: 1996-05-09  
 ; EARLIER APPLICATION NUMBER: GB 9619924.5  
 ; EARLIER FILING DATE: 1996-09-24  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Lycopersicon esculentum  
 ; US-09-180-439-4

Query Match 14.3%; Score 884; DB 4; Length 968;  
 Best Local Similarity 29.6%; Pred. No. 5e-56;  
 Matches 296; Conservative 151; Mismatches 357; Indels 196; Gaps 37;  
 QY 2 KTFSSFFLVTTLFFPSFSLSFQASPSQSLYREIHQLISFQVLPDKA--LLPDMSSNK 59  
 Db 7 KVVSS--LQFTVFY--LFTVAPAST-----EEATALKWKATFQNNNSFLASTVTS 56  
 QY 60 NPC-TFGVTCGRDKYTSIDLSKPLNVGSAVSSSL-----SLGLSLFLNSHNG 113  
 Db 57 NACKDWGVVCLNGRVNT-----LNTINAVIGTLYAPFSSLSFLENLDSNNNISG 109  
 QY 114 SV-SGFKCSALTSIDLSRNSLSPVTTLSGSCGLKFLNVSNTLD--FPKVSQGL 170  
 Db 110 TIPEIGANTLVLDINTQISG--TIPIQISLAKLQIRIFNNHNGFIPEEIG--- 164  
 QY 171 KNSLEVLIDANSISGA-----NVGWSLSDGCGELK---HLAISG 209  
 Db 165 VLSLTKLISGIFNLSGISIPASLGNMTNLSFLYENQSLGFTPEEIGYLRSLTKLSLDI 224

QY 210 NKISGDVDS--RCVNLEFLDVSSNNPSTGIP-FLGDCSALQHLIDSGNKLSGDFSRRAIS 266  
 Db 225 NELSSTIPASLGNLNNISFLYLNQLSGSIPEEIGYLSLTLSGIFNLSGISIPASIG 284  
 QY 267 TCTELKTLNITSNQFVPIPL--PLKSLQYLSIAENKFTGEIIPDLFSACDPTLTGLDLS 324  
 Db 285 NNNLSRLDLYNNKLSGISIPEEIGYLRSLTYLDGEMALNGSIPSSL--GNLNNISRLDLY 343  
 QY 325 GNHFYGAVPFPGSCSLLESALSSNNPFGELPMDTLLKRGKIVLDLSFNEPSEGEIPES 384  
 Db 344 NNKLSGISIPEEIGYLRSLTYLDGEMALNGSIPA--SIGNLNNTFMLYLNQLSGISIPEE 402  
 QY 385 LTNLSASLTLDSSNNPSCP1--LPMLCMPXNTLOEYLQNNNGFTGKIPTLSNCSB 441  
 Db 403 IGYLS-SLTELXNLNLSNGSIPASLGNL-----NLMFMLYLNQLSGISIPEEIGYLS 456  
 QY 442 LVSLHAFNYLSGTIPSSLSGLSKLRDKLMLNMBEIGLOELMYVVTLETLIDFENDLT 501  
 Db 457 LTELFLGNNSINGSIPASLGNLNNISRLYLNQLSGISIPASFGNMENLOTFLISDNDLI 516  
 QY 502 GRIPSGLSNCTNNIMNISNNRLGGEIPKIGRLNLAILELKSNNSPSGNIPDELGDCRS 561  
 Db 517 GEIPFVCNLTSLVLYVSRNNLKGVPOCJGNSIDHLIISMSNSFRGELPSISNITS 576  
 QY 562 LTMLDLNTLNFNGTIP-----AAMF-----KQSKIAANFTAGKRYVI----- 600  
 Db 577 LKILDPGRNNEGALPOFPGNISSLOVFDQNNKLSGTLPTNFSIGSLSLNLHGNELA 636  
 QY 601 -----KNDGMRK-----ECHG-----AGNLEF 618  
 Db 637 DEIPRLDNCKKQVLDLGNQNDTFFPMMLGTLPELRVLRILSNLHPISSGAEIMF 696  
 QY 619 QGIRSEQLNR--LSTRNPNITSRYVGGHT-----SPTFDN--NGSNM----- 657  
 Db 697 PDRLITDLSNNAVSQDLPSTLFEHLKGMRTVDKMEEPSEYSDSVVVVTKGLELEIV 756  
 QY 658 -----FLDMSTYMLSGTIPKEIGSMPIYFIINLGHNDISGSIPEVGDRLGNTLIDLS 710  
 Db 757 RLISLTYITIDLSNKPFGHILPSVLGDLAIRVLNVSNNALQGYIPSSLSLSTLSLIDLS 816  
 QY 711 SNKLDGRIPQMSALTMLEIDLSNNLSGPIPEMGQFETFPAPKFLNPGLCGYPLPR- 769  
 Db 817 FNQLSGEIPQOLASLTLEVLNLSHNLYLGGCIPGGPQFRTFESNSYEGNDGLRGYPVSKG 876  
 QY 770 C--DPSNADGYA--HHQSHGR-----RPASLAGSVAMGLFSPFCIFGLIVG----- 814  
 Db 877 CGKDPVSEKNYTVGALDEQESNEFEFNDFWKALMG--YGSGLCIGISIIYILISTGNLRW 935  
 QY 815 -----REMRKRRKKEALEMYAEHGNSGDRTANN 845  
 Db 936 LARIIELEHKIIVORRKKQ-----RQQRNRRRN 966

RESULT 15  
 US-08-475-891A-2  
 ; Sequence 2, Application US/08475891A  
 ; Patent No. 5859339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mang, Pamela C.  
 ; APPLICANT: Song, Men-Yuang  
 ; APPLICANT: Wang, Guo-Liang  
 ; TITLE OF INVENTION: Procedures and Materials for Confering  
 ; TITLE OF INVENTION: Disease Resistance in Plants  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Center, Eighth Floor  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475.891A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/373,375
FILING DATE: 17-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baerlan, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 02370-058910US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1023 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..1023
OTHER INFORMATION: /note="Xa21 Xanthomonas spp."
OTHER INFORMATION: disease resistance polypeptide RRR-F
OTHER INFORMATION: from rice (Oryza sativa)"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1010
OTHER INFORMATION: /note="Xaa = Leu when position
OTHER INFORMATION: 5471 of RRR-F = G or Phe when position
OTHER INFORMATION: 5471 of RRR-F = C"
US-08-475-891A-2.

```

```

Query Match 14.2% Score 875.5; DB 2; Length 1023;
Best local Similarity 27.6%; Pred. No. 2,3e-55;
Matches 321; Conservative 166; Mismatches 404; Indels 273; Gaps 45;

```

```

QY 89 SAVSSLSITGLESFL--SNHNGSVSGFKCSAITSITLSDSRNSISG-PVTTLTSLG 145
DB 7 SVWISSLL--LLIGPASDDDAALAAARTSTGVAAATNSCSLSHPCYT--RG 58
QY 146 SCGGKLFVAVSSN--TLDPKGYSGGLKINSLEVLDSANSISGANVAVVYDSCGE-- 201
DB 59 ACT-LASWNTSGHGHCTWGVVCGRRARHPRVYKLLRS--SNLGG-IISPSLNTLS 113
QY 202 -LKHLSISGKISGVDVSRGVNLEFLDVSSNNPSTGIPFLDCSALQHDISGNKLSGD 260
DB 114 FLRELDLSNYSIGEIP-----PELSRLSRQLLEISGNSIQGS 152
QY 261 FSRALSTTELKLNLISSNQFVPIPLK-SLOYLS---LAENKP---TGEI--PDF 310
DB 153 IHAAGACTKLTSLDLSHQ---LRVPAETSLFVPSHQMLCOERPHLLGNLTTPSV 208
QY 311 LSGACDTLT-----GLD--LSGNHFYGAVPFPFGSCSLLESLAL--SSNFS 353
DB 209 FDLTCNRLRSRYTFARAQQOSIDYEFCAINNLTMGMPNSIWNLSLAASFCAKRYKLG 268
QY 354 GELPMDTLLKMGKLVLLSPNEFSGELPESLTLNLSASLTLTLDSSNNSFGPIIPNLCON 413
DB 269 GMIPFNARKTLLHLEVVVGTRFRGKIPASVANAS--HLTRLQIDGNLFGSIIITSGF--G 325
QY 414 PKNLTQELIYONGFTGK-----IPPTLNSCSSEIVSIHLSPNYLSGTIPSSLSGLSKLR 467
DB 326 RLRLNLTLYLWRNLFQTRQEDMGWGISDLTNCSTKQTLIDGENNLGSGVLPNSPNSLS 382
QY 468 DUKMLNMLEGIPQELMYVKTLETLIDFNDLTGEIPSGLSNCTNLNWSLSNNRLTGE 527
DB 383 -----TSLSFLALDLNKITGSIPIKDIQNLIGLQHLXLCNNPFGS 422

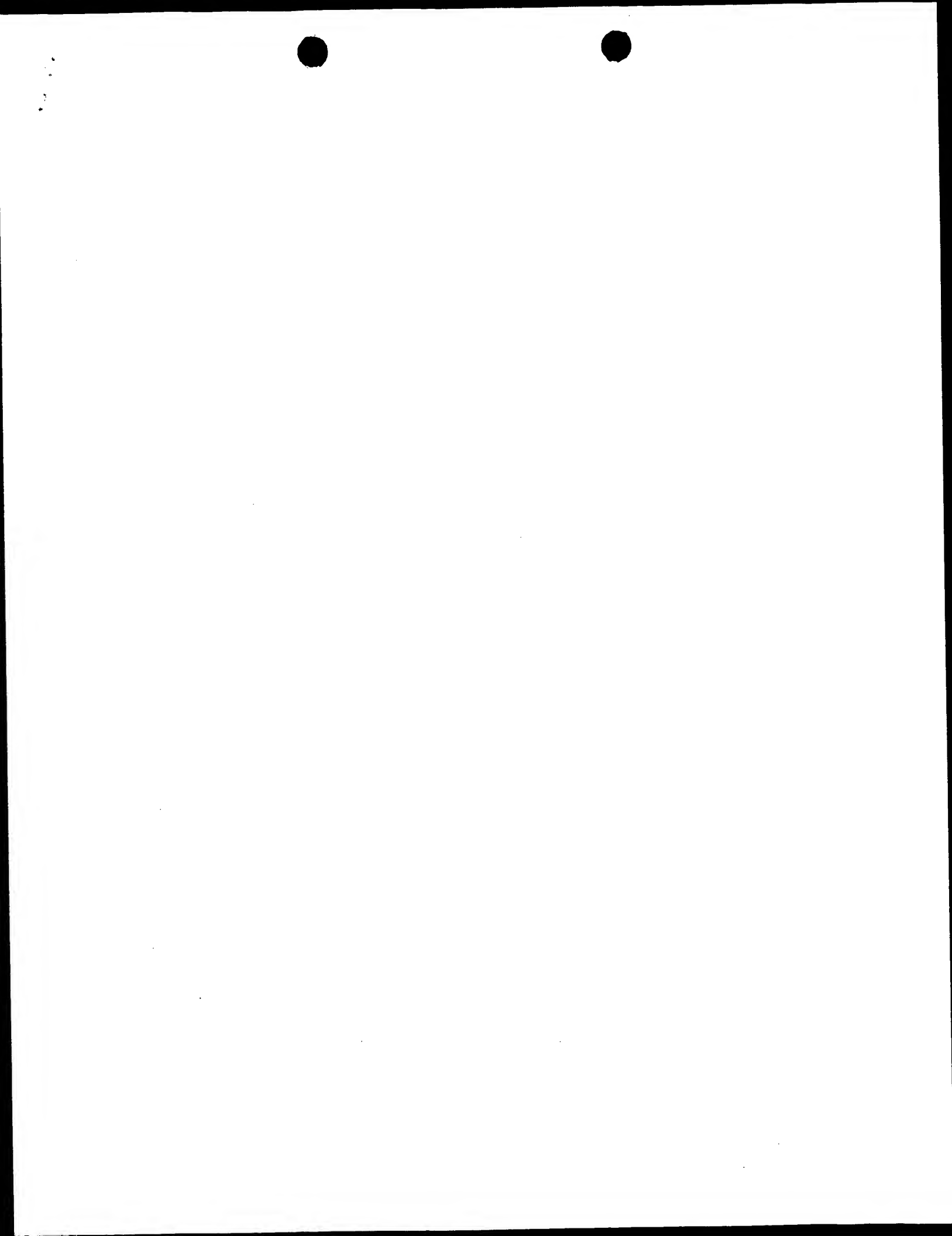
```

```

QY 528 IPKNIIGLEMLAILKLSNNSPSGNIPELGDCSRLLIMDLTNLFGFTIPAMFKQSGKI 587
DB 423 LPSLSIGRLRNGLIVAVENNLSSIPALGNLLEMLNLLGTRKSGWLPYTLNLTNLL 482
QY 588 AANFIAGKRYVYIKNDGKKECHAGNLEFQGISRSEQNLRLSTNPNCTITSRYVGHTS 647
DB 483 SL-----GLSHLAPQSGLDL-----YTSOPHNCVIAFTIISGSGOI 517
QY 648 PTFD-----NNGSMFLDMSYNNLSGY-IPKETSMPYFIINLGHNDISGISPEVGL 701
DB 518 PQVTLQIVQTPKKNITVSKNTLGGIRYPOETGHLKNIVEQNRIDIS-KIPNTLQDC 576
QY 702 RGNILDLSSNKLDPRIQAMSAITLLEIDLSNNNLSGPIPEMGQFEPFPAKFLNN-- 759
DB 577 QLRVLYLQNNLLSGSIPALGQKLFTLDSNNLSQITP--SLRRYYASFLEPF 634
QY 760 PGLCGYPLPCDPSNADGYAHH---QRSHGRPPASLAGSVAMGLFSPVCI FGLILVGR 815
DB 635 QQLCG-----GSANHCVLSSQMHGSGSKAMPNSVVEYLI--YICLDVPELL-- 678
QY 816 EMRK-----RRKKAELMELMAEGHSGSGDRTANNTN 847
DB 679 ENRKHPALPISVSLVALAILSSLYLLITMNRKTKGAPRTSKWGH----- 726
QY 848 WKLTGVKALSLINLAFFKPLRKLTFADLLQATNGFNHDSLIISGSGFGDYKALIKDGA 907
DB 727 -----PL--VSYPOLVYATDGFAPTNLGSGSFSVYKRLLENPKA 765
QY 908 VAIKLIHVSQGDREPMAMETIKIRKRLVPLLGYCKYGER-----LVNBEVKYG 962
DB 766 L-----KSFPAECBALNMRHRLVKLVITCSSIDNRGNDFKAIYDPMNG 812
QY 963 SLSDVQ---DPKGGVTLKSTRKKAIGSARGIAFLHNCSPHIIRDMKSSNVLLDE 1019
DB 813 SLEDWHPETNDQADQRLNLRHRYTLLDVAICALDYLRHGPPEPVVHCDVKSNNVLLDS 872
QY 1020 NLEARVSDFGMARLM-----SAMDTLSVSTLAGTGPVVPPEYQSFRCSTKGDVYSGVV 1075
DB 873 DVAIVHVGSGGLARILVDGSLIQOSTSMGFRGTTGYAABYGVGHIASTHGDIYSYL 932
QY 1076 LLELITGRPTDS---PDGNNILVGVKQHAU-----RLSDVPEPELM-----K 1118
DB 933 VVEIYTGKRPDSTSRPDIG-----LRQYVELGHRVTDVDTLLIDSEWNLNSTN 985
QY 1119 EDPALER--LLQHLKAVACLDD 1140
DB 986 NSPCRRTICIVSLRLGLSCSD 1009

```

Search completed: March 10, 2003, 18:19:21  
Job time : 33.6693 secs



	RESULT 1
AAW97819	
ID	AAW97819 standard; Protein; 1196 AA.
XX	
AC	AAW97819;
XX	
DT	21-MAY-1999 (first entry)
XX	
DE	Arabidopsis steroid receptor Binl.
XX	
KM	BIN1; steroid receptor; receptor kinase; transgenic plant; brassinosteroid; disease resistance; crop protection; contraceptive. KM XX OS Arabidopsis thaliana. XX FN W09859039-A1. PD 30-DEC-1998. PF 24-JUN-1998; 98WO-US31100. PR 24-JUN-1997; 97US-0861706. PA (SALK ) SALK INST BIOLOGICAL STUDIES. PI Chory J, Li J; DR WPI; 1999-081275/07. N-PSDB; AAX07356. PT New receptor kinase Binl involved in brassinolide signalling - useful for promoting increased yield and disease resistance in

plants and for modulating oocyte maturation  
claim 3, page 52, 72pp; English.

This is the amino acid sequence of a novel plant steroid receptor kinase, designated Bin1, which is involved in the pathway for the synthesis of the plant steroid hormone, brassinolide. 18 New Arabidopsis dwarf mutants were identified that lacked the ability to respond to brassinolide, and were named bin mutants. The bin1 mutations were used to map the gene to a small interval on Arabidopsis chromosome 4. Bin1 was cloned by map-based cloning. A Bin1 polynucleotide (see AAX07356) was identified within this interval by sequencing the wild-type and mutant alleles of this nuclear acid. Overexpression of Bin1 in transgenic plants provides plants characterized as having enhanced biomass and increased seed yield. Bin1 expression may also increase resistance to pesticides. Inhibition of Bin1, e.g. using antisense oligonucleotides (AON), is used to render plants male-sterile, and to reduce their stature or yield, e.g. for creating dwarf varieties. Since Bin1 homologues may be involved in regulation of the menstrual cycle and uterine function, Bin1, antibodies and AON may be useful as contraceptives, for improving success of in vitro fertilisation and to prevent premature labour. Transgenic animals are also provided, and are models for studying steroid-receptor interactions or can be used to screen for therapeutic agents.

Sequence 1196 AA:

Query Match 100.0%; Score 6183; DB 20; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKTSSFFLSTVTLTFFSFFSLSFQASPSQSLYREIHOLISFKVLPDKNLLPDMSSNKN 60  
1 MKTSSFFLSTVTLTFFSFFSLSFQASPSQSLYREIHOLISFKVLPDKNLLPDMSSNKN 60  
1 PCTPDGVTCTRDDKVTISLSSKPLNVSFAVSSLSLTGLESFLNSHINSVSGKCC 120  
1 PCTPDGVTCTRDDKVTISLSSKPLNVSFAVSSLSLTGLESFLNSHINSVSGKCC 120  
61 PCTPDGVTCTRDDKVTISLSSKPLNVSFAVSSLSLTGLESFLNSHINSVSGKCC 120  
121 SASLTSLSLRSNLSGAPVTLTSLGSGGLKPLNVSNTLPDGVSGGLKLNLSLEVDL 180  
121 SASLTSLSLRSNLSGAPVTLTSLGSGGLKPLNVSNTLPDGVSGGLKLNLSLEVDL 180  
121 SASLTSLSLRSNLSGAPVTLTSLGSGGLKPLNVSNTLPDGVSGGLKLNLSLEVDL 180  
181 SANSISGANVGVMLSDGCGELKHLAISGNKISGDVDSRCVNLLEFLDVSSNNFSTGIPF 240  
181 SANSISGANVGVMLSDGCGELKHLAISGNKISGDVDSRCVNLLEFLDVSSNNFSTGIPF 240  
241 LDDCSALQHLDISGNKLSGDFSRALSTCTELKLNINSSNQPGPIPLPKLSIQLSLAE 300  
241 LDDCSALQHLDISGNKLSGDFSRALSTCTELKLNINSSNQPGPIPLPKLSIQLSLAE 300  
301 NKFTGEIPPLFGACDTLTGLDLGNNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
301 NKFTGEIPPLFGACDTLTGLDLGNNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
361 LKMRGLKVLDSFNFSGELPESLTNLSASLTLDLSSNNFSGPIPLPMLCONPKTLOE 420  
361 LKMRGLKVLDSFNFSGELPESLTNLSASLTLDLSSNNFSGPIPLPMLCONPKTLOE 420  
421 LYQNNNGFGKIPPLTNSGSELVSLHSFNYSGLTSSLSGLSKLRDLKMLNMLEGI 480  
421 LYQNNNGFGKIPPLTNSGSELVSLHSFNYSGLTSSLSGLSKLRDLKMLNMLEGI 480  
481 POELMTVKTLETLIDPNDLTGEIPGLSNCTVLMWISLNNRLTGEIPKWIQRLENTAI 540  
481 POELMTVKTLETLIDPNDLTGEIPGLSNCTVLMWISLNNRLTGEIPKWIQRLENTAI 540  
541 LKLSNNSFSGNIPDELGDGCSLIWLDLNTNLNGTTPAAMFKOSGKIAPNAGKRYVYI 600  
541 LKLSNNSFSGNIPDELGDGCSLIWLDLNTNLNGTTPAAMFKOSGKIAPNAGKRYVYI 600  
541 LKLSNNSFSGNIPDELGDGCSLIWLDLNTNLNGTTPAAMFKOSGKIAPNAGKRYVYI 600

601 KNDGKKCHGAGNLEFFQGRBSQNLRLSTRMPCNITSRVYGHSPEDNNNGNMFLD 660  
601 KNDGKKCHGAGNLEFFQGRBSQNLRLSTRMPCNITSRVYGHSPEDNNNGNMFLD 660  
661 MSYMLSGYIPKEIGSMPLYFIINLGHNDISGSIPEDEVDLRGLNLDSSNKLDRIPQ 720  
661 MSYMLSGYIPKEIGSMPLYFIINLGHNDISGSIPEDEVDLRGLNLDSSNKLDRIPQ 720  
721 AMSALTMLTEIDLSNNLSGPIPEMGQFETFPKPLNNPGLCGYPLRCDPSNADGYAH 780  
721 AMSALTMLTEIDLSNNLSGPIPEMGQFETFPKPLNNPGLCGYPLRCDPSNADGYAH 780  
781 HORSGRBPASLAVMGLIFSFCVIFGLIVGEREKRRKKEALEMYABEGHNSGD 840  
781 HORSGRBPASLAVMGLIFSFCVIFGLIVGEREKRRKKEALEMYABEGHNSGD 840  
841 RTANNTNKLTVGEKALSTINLAPEKPLRKTTPADLLQATNGFNNDSLISGGRGDYKA 900  
841 RTANNTNKLTVGEKALSTINLAPEKPLRKTTPADLLQATNGFNNDSLISGGRGDYKA 900  
901 IIKQSAVAIAIKKLIVHSGQGRREMAEMETIGKIKHNLVPLGYCKVGERLLVNEVMK 960  
901 IIKQSAVAIAIKKLIVHSGQGRREMAEMETIGKIKHNLVPLGYCKVGERLLVNEVMK 960  
961 YGSLBDVLQDPKKGAVKLLSTRKRIAIGSARGLAFHHNCSPHIHRDMKSNVLLDEN 1020  
961 YGSLBDVLQDPKKGAVKLLSTRKRIAIGSARGLAFHHNCSPHIHRDMKSNVLLDEN 1020  
1021 LEARYSDFGMARLMSAMDTLHLSVSTLAGTGYVPEYVQSFRCSTKGDVYGVLLLEL 1080  
1021 LEARYSDFGMARLMSAMDTLHLSVSTLAGTGYVPEYVQSFRCSTKGDVYGVLLLEL 1080  
1081 TGRKRPDSDPDGNNLVGVVQKALRISDVDPPELMKEDPALTEILLQHKVAVACLD 1140  
1081 TGRKRPDSDPDGNNLVGVVQKALRISDVDPPELMKEDPALTEILLQHKVAVACLD 1140  
1141 RAMRRTVQVWAMFEKIEIAGSGIDSQSTIRSIDGFSSTIBMDMSIKVEYPEGKL 1196  
1141 RAMRRTVQVWAMFEKIEIAGSGIDSQSTIRSIDGFSSTIBMDMSIKVEYPEGKL 1196

RESULT 2  
AAB67443  
ID AAB67443 standard; Protein; 1196 AA.  
XX  
AC AAB67443;  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Amino acid sequence of an Arabidopsis Br1 protein.  
XX  
KW Disease resistance protein; Xa21; RKK gene; transgenic plant;  
XX Xanthomonas; plant pathogen; Br1 protein; RCH10 protein.  
XX  
OS Arabidopsis sp.  
XX  
PN W0200109283-A2.  
XX  
PD 08-FEB-2001.  
XX  
PF 28-JUL-2000; 2000MO-US20714.  
XX  
PR 28-JUL-1999; 99US-036313.  
XX  
PA (REGC) UNIV CALIFORNIA.  
XX (SHUK) SALK INST BIOLOGICAL STUDIES.  
XX Ronald P. He Z, Chory J, Lamb C, Li J;  
XX WPI; 2001-159858/16.  
XX DR N-PSDB; AAF54982.  
XX  
PT Chimeric plant receptors comprising a polynucleotide encoding a RKK

receptor containing a heterologous extracellular domain and a kinase domain from a Xa21 polypeptide, useful for modulating plant responses to pathogens

Disclosure; Page 38-42; 47pp; English.

The present sequence represents a Br11 protein. The specification describes chimeric receptors which are used for modulating plant responses to pathogens. The receptors comprise a heterologous extracellular domain (e.g. from a Br11 protein or Rch10 protein) and a kinase domain (e.g. from disease resistance protein Xa21). The Xa21 gene is a member of a disease resistance gene referred to as RKK genes. When Xa21 is present in a transgenic plant, it confers resistance to Xanthomonas spp.. The plant receptors containing heterologous domains are useful for modulating plant responses to pathogens including viruses, bacteria, nematodes, fungi or insects. The nucleic acids can be used to confer desired traits on essentially any plant.

Sequence 1196 AA;

Best Match 99.3%; Score 6140; DB 22; Length 1196;

Best Local Similarity 99.4%; Pred. No. 0; Matches 1193; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 1 MKTFSSPFLVTLTLPFSEFSLSFQASPSQSLYREIHQILSFQVLPDKMLPDMSSNKK 60
DB 1 MKTFSSPFLVTLTLPFSEFSLSFQASPSQSLYREIHQILSFQVLPDKMLPDMSSNKK 60
QY 61 PCTFDVTCDDKVTSLDSSKPLNVFSAVSSSLSLTGLTSLFSLNHSINGSVGFC 120
DB 61 PCTFDVTCDDKVTSLDSSKPLNVFSAVSSSLSLTGLTSLFSLNHSINGSVGFC 120
QY 121 SABLTLSDLSRNLGSGPVTTLTSLGSCSGLKPLNVSSNTLDPFGKXSGGLKNSLEVL 180
DB 121 SABLTLSDLSRNLGSGPVTTLTSLGSCSGLKPLNVSSNTLDPFGKXSGGLKNSLEVL 180
QY 181 SANSISGANVGVVLSGGGCELKHLAISGKISGVDVSRCVNLEFLDVSSNNFSGTIPF 240
DB 181 SANSISGANVGVVLSGGGCELKHLAISGKISGVDVSRCVNLEFLDVSSNNFSGTIPF 240
QY 241 LGDCSALQHLIDISGNKLSGDFSRALSTCTELKLNISNOFVPIPLPKSLQVYSLAE 300
DB 241 LGDCSALQHLIDISGNKLSGDFSRALSTCTELKLNISNOFVPIPLPKSLQVYSLAE 300
QY 301 NKRTGIIIPPLSACCTITGLDISGHPFGAVPPFGSCSLLESLALSSNNFSGELPMDT 360
DB 301 NKRTGIIIPPLSACCTITGLDISGHPFGAVPPFGSCSLLESLALSSNNFSGELPMDT 360
QY 361 LKMRGLKYLDSFNFSGELPESLTNLASLTLTLDSSNNFSGPLTPVLCONPKVITOE 420
DB 361 LKMRGLKYLDSFNFSGELPESLTNLASLTLTLDSSNNFSGPLTPVLCONPKVITOE 420
QY 421 LYLQNNGTGKIIPPLTNSCSEIVSLHSFNVLSTGTPSSLSGLSKLRDKMLNLEGEI 480
DB 421 LYLQNNGTGKIIPPLTNSCSEIVSLHSFNVLSTGTPSSLSGLSKLRDKMLNLEGEI 480
QY 481 POELMAYVTKLETLIDRNLTGRIPSGLSCTNLMWISLNNRLTGEIIPKWIIGLENTAI 540
DB 481 POELMAYVTKLETLIDRNLTGRIPSGLSCTNLMWISLNNRLTGEIIPKWIIGLENTAI 540
QY 541 LKLSNNSFSGNIPDELGDRSLIWLDMTNLNFNGTIPAMFQSGKIANFAGRRYYI 600
DB 541 LKLSNNSFSGNIPDELGDRSLIWLDMTNLNFNGTIPAMFQSGKIANFAGRRYYI 600
QY 601 KNDGKMKCECHGAGNLEFQIRSEQLNRLSTNPNCTISRYVGGTSPFPDNGSMFLD 660
DB 601 KNDGKMKCECHGAGNLEFQIRSEQLNRLSTNPNCTISRYVGGTSPFPDNGSMFLD 660
QY 661 MSYNMLSGYIPKISGMPYFLINLGHNDISGIPDEVGDLKGLNITLSSKLDGRIPQ 720
DB 661 MSYNMLSGYIPKISGMPYFLINLGHNDISGIPDEVGDLKGLNITLSSKLDGRIPQ 720
QY 721 AMSALTMLTEIDLSSNNLSGPIPEMGOETFPFAKFLNPGLCGYPLPCDPSNADGVAH 780

```

```

DB 721 AMSALTMLTEIDLSSNNLSGPIPEMGOETFPFAKFLNPGLCGYPLPCDPSNADGVAH 780
QY 781 HQRSHGRBPASIASGVAMGLLPSPVCFIQLILVGRBMRKRRKKEALEMYAEHGNSGD 840
DB 781 HQRSHGRBPASIASGVAMGLLPSPVCFIQLILVGRBMRKRRKKEALEMYAEHGNSGD 840
QY 841 FRANTTNMKTGVKALSLINLAFFKPKRLTPADLQATNGFHNDSLISGGGADYVKA 900
DB 841 FRANTTNMKTGVKALSLINLAFFKPKRLTPADLQATNGFHNDSLISGGGADYVKA 900
QY 901 LKDSAAVAIKLIVNSQDREFMAEMETIGIKHRLVPLLGCKVGDRLLVNEMK 960
DB 901 LKDSAAVAIKLIVNSQDREFMAEMETIGIKHRLVPLLGCKVGDRLLVNEMK 960
QY 961 YGSLDEVADPCKGVKULSTRKRIAGARGIAFLHNCSPHIIHRDKSSNVLLDEN 1020
DB 961 YGSLDEVADPCKGVKULSTRKRIAGARGIAFLHNCSPHIIHRDKSSNVLLDEN 1020
QY 1021 LEARVSDFGMARLMSAMDTLHVSSTLAGTPGVPPPEYOSFRCTGDVYSYGVVLELL 1080
DB 1021 LEARVSDFGMARLMSAMDTLHVSSTLAGTPGVPPPEYOSFRCTGDVYSYGVVLELL 1080
QY 1081 TGRKPTDSPDFGDNVLVGWVQKAKRLISDVPELMEKEDPALEIELLOHLKVAACLDD 1140
DB 1081 TGRKPTDSPDFGDNVLVGWVQKAKRLISDVPELMEKEDPALEIELLOHLKVAACLDD 1140
QY 1141 RAMRRPTVQVMAFKEIQAQSGIDSGTIRSIDGGFTIEMWMSIKEVPEGL 1196
DB 1141 RAMRRPTVQVMAFKEIQAQSGIDSGTIRSIDGGFTIEMWMSIKEVPEGL 1196

RESULT 3
AAE19490
ID AAE19490 standard; Protein; 1196 AA.
AC AAE19490;
DT 31-MAY-2002 (first entry)
XX
DE Brassinosteroid receptor protein encoded by Br11 DNA.
XX
KW Plant receptor protein; disease resistance protein; plant development;
KW RRR protein; disease resistance; cell free assay; gene therapy;
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Misc-difference 162
FT /note= "Encoded by TTT"
XX
XX W0200210367-A1.
XX
XX 07-FEB-2002.
XX
XX 28-JUL-2000; 2000WC-US20604.
XX
XX 28-JUL-2000; 2000WC-US20604.
XX
XX (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
XX Chory J, Lamb C, He Z;
XX
XX WPI; 2002-227084/28.
XX
XX N-PSDB; AAD30800.
XX
PT Heterologous polynucleotide encoding chimeric plant receptors for
PT controlling plant development and disease resistance, has leucine-rich
PT repeat domain, transmembrane domain, and cytoplasmic protein kinase
PT domain
XX
PS Disclosure; Page 49-50; 54pp; English.

```

XX The present invention relates to heterologous nucleic acid molecules  
 CC encoding chimeric plant receptor proteins comprising a leucine-rich  
 CC repeat (LRR) domain, a transmembrane domain and a kinase domain from  
 CC cytoplasmic RRR (disease resistance proteins) proteins such as Xa21.  
 CC The nucleic acid sequences are useful for identifying ligands for  
 CC receptor or receptor-like kinase. The chimeric receptors are useful  
 CC for controlling plant development and/or disease resistance. They are  
 CC used in cell free assay useful for determining the ability of a test  
 CC compound to bind to or modulate the activity or expression of the  
 CC receptor. Sequences of the invention are also used in gene therapy.  
 CC The present sequence is brassinosteroid receptor protein encoded by  
 CC brr1 DNA. This sequence is used in the invention.  
 XX  
 SQ Sequence 1196 AA;  
 Query Match 99.1%; Score 6130; DB 23; Length 1196;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 1188; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 1 MKTSSFLSYTLTFFSFLSFGASPSGLYREIHQILSFQKVLDPKLLPWSNNKN 60  
 1 MKTSSFLSYTLTFFSFLSFGASPSGLYREIHQILSFQKVLDPKLLPWSNNKN 60  
 61 PCTPGVTCRDKKTSIDLSKPLNWFSAVSSLSLTGLBSLFLSNHSHNGSVGPKC 120  
 61 PCTPGVTCRDKKTSIDLSKPLNWFSAVSSLSLTGLBSLFLSNHSHNGSVGPKC 120  
 121 SASLISLDSRSLSGPYTTLTSLGSCGLKFLNVSNTLDFPGKVSGLKLNLSLEVL 180  
 121 SASLISLDSRSLSGPYTTLTSLGSCGLKFLNVSNTLDFPGKVSGLKLNLSLEVL 180  
 181 SANSISGANVGVWVSDGCGELKHLAISGKISGDVDRSCVNLFLDVSNNSTGIPF 240  
 181 SANSISGANVGVWVSDGCGELKHLAISGKISGDVDRSCVNLFLDVSNNSTGIPF 240  
 241 LGDCSALQHLIDISGNKISGDFSRATSTCTELKLNINSSNOFGPIPLPKSLQTLIAE 300  
 241 LGDCSALQHLIDISGNKISGDFSRATSTCTELKLNINSSNOFGPIPLPKSLQTLIAE 300  
 301 NKFTGEIPPLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
 301 NKFTGEIPPLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
 361 LKMRGLKTLDSFNFSGELPESLTNLASLITLDSNNNSGPIPLCONPKNTLOE 420  
 361 LKMRGLKTLDSFNFSGELPESLTNLASLITLDSNNNSGPIPLCONPKNTLOE 420  
 421 LKMRGLKTLDSFNFSGELPESLTNLASLITLDSNNNSGPIPLCONPKNTLOE 420  
 421 LKMRGLKTLDSFNFSGELPESLTNLASLITLDSNNNSGPIPLCONPKNTLOE 420  
 481 POELMTVKTLETLIDPNDLTGEIPSGLSNCTNLMNLSNNRLTGEIPKWI GLENLAI 540  
 481 POELMTVKTLETLIDPNDLTGEIPSGLSNCTNLMNLSNNRLTGEIPKWI GLENLAI 540  
 541 LKISNNSPSGNIPDELGDGCSLITWLDLNTNLNNGTIPAMFROSGLIANFAGKRYVI 600  
 541 LKISNNSPSGNIPDELGDGCSLITWLDLNTNLNNGTIPAMFROSGLIANFAGKRYVI 600  
 601 KNDGKKECHGAGNLLEFQGISREQLNRLSTRNPNCTISVYGHTSPFPDNGSMFLD 660  
 601 KNDGKKECHGAGNLLEFQGISREQLNRLSTRNPNCTISVYGHTSPFPDNGSMFLD 660  
 661 MSYNMLSGYIPKEIGSMPIYFIINLGNDISGSIPEVGDRLKINLIDSSNKLDRIRIQ 720  
 661 MSYNMLSGYIPKEIGSMPIYFIINLGNDISGSIPEVGDRLKINLIDSSNKLDRIRIQ 720  
 721 AMSALMTLTDIDSSNNLSGPIPEMGOFETPPAKFLANNGLGYPILPCDPENALAGYAH 780  
 721 AMSALMTLTDIDSSNNLSGPIPEMGOFETPPAKFLANNGLGYPILPCDPENALAGYAH 780  
 781 HORSRRPAPSLAGSVAMGLLFSFVCIFGLILVGRMRKRRKKEALEMYAEHGNSGD 840

DB 761 HORSRRPAPSLAGSVAMGLLFSFVCIFGLILVGRMRKRRKKEALEMYAEHGNSGD 840  
 QY 841 RTANNTNKLTVGEALSTINLAEPKRLKLTENDLOANTNGFNDSLIGSGGVDYKA 900  
 DB 841 RTANNTNKLTVGEALSTINLAEPKRLKLTENDLOANTNGFNDSLIGSGGVDYKA 900  
 QY 901 LKDSAAVAIKKLHVSGQGRFPAEMETIGIKHNLVPLGYCKVGERLLVNEVMK 960  
 DB 901 LKDSAAVAIKKLHVSGQGRFPAEMETIGIKHNLVPLGYCKVGERLLVNEVMK 960  
 QY 961 YGSLDVLQDPKKGKVKLSTTRKRIAGSARGLAFIHNCSPIIHRDMKSSVLLDEN 1020  
 DB 961 YGSLDVLQDPKKGKVKLSTTRKRIAGSARGLAFIHNCSPIIHRDMKSSVLLDEN 1020  
 QY 1021 LEARVSDGMAKLMANTHLSYSTLAQTPGVPEYQSFRCSTKDVYGVVLELL 1080  
 DB 1021 LEARVSDGMAKLMANTHLSYSTLAQTPGVPEYQSFRCSTKDVYGVVLELL 1080  
 QY 1081 TGRPTDSDPDGNNLVGVVQKAKLISDVDPPELMKEDPALEIELLOHLKVAACLD 1140  
 DB 1081 TGRPTDSDPDGNNLVGVVQKAKLISDVDPPELMKEDPALEIELLOHLKVAACLD 1140  
 QY 1141 RAMRPTVQVAMFEKIOAGSGIDSGTINSIEDGFTIEMVDMSTKEYPEGL 1196  
 DB 1141 RAMRPTVQVAMFEKIOAGSGIDSGTINSIEDGFTIEMVDMSTKEYPEGL 1196

RESULT 4  
 AAG79244  
 ID AAG79244 strand; Protein; 1121 AA.

AC AAG79244;  
 DT 03-JAN-2002 (first entry)

DE Amino acid sequence of a gene concerning brassinosteroid sensitivity.  
 XX OsBR1; brassinosteroid sensitivity; d61 locus; rice;  
 KW internode elongation; internode cell; lamina joint.

OS Oryza sativa.  
 XX WO200173036-A1.

PN 04-OCT-2001.

XX 30-MAR-2001; 2001WO-UP02770.

XX 31-MAR-2000; 2000JP-0101276.

XX (NAG-) NAT INST AGROBIOLOGICAL SCI.

PA Tanaka H, Kayano T, Matsunaka M;

XX WPI; 2001-616505/71.

XX N-PSDB; AAI65842.

PT Gene relating to brassinosteroid-sensitivity of plants, useful in  
 PT controlling growth and development of transformants including rice to  
 PT improve harvest and crop yield for animal feed or dwarfism to enhance  
 PT ornamental effect -

XX Claim 1; Page 72-80; 87pp; Japanese.

XX The present sequence is encoded by a gene, designated OsBR1, which  
 CC encodes the brassinosteroid sensitivity of plants. The OsBR1 gene is  
 CC located to the d61 locus. The gene is involved in the growth and  
 CC development of rice, for example, internode elongation via the induction  
 CC of the elongation of internode cells and bending of lamina joints.  
 CC The OsBR1 gene is useful in controlling growth and development of  
 CC transformants. Transformant rice plants can be obtained to improve  
 CC harvest and crop yield for providing more animal feed in agriculture,



CC or plants with dwarfism can be produced to enhance ornamental effect  
CC and added value in horticulture and related industries.  
XX

SQ Sequence 1121 AA;

Query Match	49.38;	Score	3050.5;	DB	22;	Length	1121;
Best Local Similarity	53.98;	Pred	No	6	88-332.		

QY 39 LI5FKYLDLPKXNLLPDMSSKNKNCCTDGVYCRDQKXTSIDLSKPLXNVGSAVSSGLSL 98  
31 LEEFROAVFQOALLKXNGSGDGA CRPPGAGCRNRRGTLSTLSLGVPLINAFRVAATLTQ 90  
QY 99 TGLESLFSLNSHINGSV--GFKCSASLTSLDLSBN-SISGVYTLTSLGS-CGSLKTL 153  
DB 91 GSVEVLTSLRGANSVALSAAGARCSKLOALDLSGNALRGSVADVAALASACGLKTL 150  
QY 154 NVBSNLTLDPPGKV--SGGLKXLSLEVLDLSANSISGANYVWYLDSCGELKHLAISGNK 211  
151 NUSGDVVG-AAKVGCGGCGGAGLJDLSDLTSLNNKTTDDSPDLRWVYVAGVAYRMJDLALNR 209  
QY 212 ISGDVDRSCVNLFEFVDSNNFSTGTIPELGDCSALQHLDISGNKLSGDF-SRAISTCTE 270  
DB 210 IS-----GVPEFTNCSGELOYLDLGNLIVGEVPGALSDCRG 246  
QY 271 LKLLNISSNOVGPPIPELPKLSLOYLSLANKFTGRIPEPLSGACDITLGLDLSGNHYG 330  
DB 247 LKVLN-----LSNHLIAG 259  
QY 331 AVPPFSGCSLLESLALSSNNFSGELPMDTLLKMGKLVLDLSFNEFSGELPELSLTNLSA 390  
DB 260 VEPPIAGLTSJNALNLSNNFSGELPEGAFAKLOQLTALSLSFHFNHSGIPDTYASL-P 318  
QY 391 SLTLTLDSSNNSGPIPLMLCONPKNTLOELYLONNGFPGKTPPLTSLNCSBELVSLHSFN 450  
DB 319 ELQOALDSSNTSGTIPSSLCDPPMSKHLTLQNNVYLTGGIPDAVSNCTSLVSLDLSLN 378  
QY 451 YLSGTIPSSLSGLSRDLKMLNLMEGIEPOLMYVYTLLETLIDPFMDLGEIPLSGLSN 510  
DB 379 YINGSIPASLGLDGLMLQDLIMQNLBEGRIPASLSRIGLEHLIIDYGLTGSIPPELAK 438  
QY 511 CYNLWISLSNNRLTGEIIPKWIGRLENTALIKLSNNSPSGNIPELDGCRSLIWLDLNTN 570  
DB 439 CTKLWISLSNRLSGPIPSMWLGKLSYALIKLSNNSSGPIPELGDQCGVLWDLNSN 498  
QY 571 LFNGTIPAMFSGKLIANFIAGRRVYIYKNDGKKECHAGNULFEQIRISQNLRLS 630  
DB 499 QUNGSIPKELAKQSKAMVGLVGPYYLLNDELISSECRKGSLEBFTSIRPDLSRMP 558  
QY 631 TRNPNCTSRVYGHTSPPTDNNGSMFLDMSVYMLSGYIPKEIGSPMYLTLNLGHNDI 690  
DB 559 SKKLGNF-RMYVSGTXYTFKNGSNMIFLDSLSTYOLDBALITGELGDMFYMLTMVGNL 617  
QY 651 SGLSPDEVGDRLGNTLIDLSNKLDMGRIPOAMSAITMLTEIDLNNMLSGPIPEMQOFET 750  
DB 618 SGLTIPSLAEAKKVALVDLSYNOLEGPIPNFSFALS-LSBINLNNQNLNGTIPELGSLAT 676  
QY 751 FPPAKFLNPGICGPIPRCDPSNADGYAHHQHGRPPASLAGSVAMGLIFSVCITFGL 810  
DB 677 FPKGYENNTGICGPIPRCHSPSSRSDNQ-SH-ROASMASSIAMGLIFSFCITVI 734  
QY 811 ILVGRNEMKRRRKKAEAL--EMVAFGHNSGSDRTAANTNNK--LTGVKELASTIYLAPEK 866  
DB 735 IIAIGSKRRRLKNEBASRSDIYIDRSHS--ATMNSDMRQNLSTG-NLJSLINLAPEK 790  
QY 867 PLRKLTPDLLOATNGRHNSLISGGFGRDYKAILKDGSAVAIKLIVHSGQDREFFMA 926  
DB 791 PLQNLTLADIVEANTGPHIACQISGGGVDYKQALDQKVAIKLIVHSGQDREFFA 850  
QY 927 EMETIGKIKHNLVPLIGYCVGDERLLNNEVMYXGSLLEVLDQPKKGYKALSTRKTI 986  
DB 851 EMETIGKIKHNLVPLIGYCVGDERLLNNEVMYXGSLLEVLDHDKRKTIGKKLWEARRKI 910

QY	387	AIGSARGLATLHNCSPIIHRMKSNNLTLEENEARSDGMRKMSADTHLSVTL	1046
Db	911	AVGARGRLATLHNCLPHIIHRMKSNNLTDEQLFARSDGMRKMSVVDTHLSVTL	970
QY	1047	AGTGPVPEPEYOSFRCSTKGDVYSYGVALLTELTKGKPTDSPDFG-DNNLTGWMYKQAH	1105
Db	971	AGTGPVPEPEYOSFRCSTKGDVYSYGVALLTELTKGKPTDSADGEDNNLVGMWKQTK	1030
QY	1106	LRISDVDPPLMEDPALTELLOHLKAVAVACLDRAMRPTMVOVMAMFEIOAGSID	1165
Db	1031	LKTTDVFDPPLLNKEDSVLELLEHLKACACLDLRPSRRPTMLKVMAMFEIOAGSTVD	1090
QY	1166	SOST---IRSIEDGSGSTIEWYMSIKVEPBEK	1195
Db	1091	SKTSSMAAGSIDGSGY---VLDMPREKBEK	1120

RESULT 5  
ABB92294

AC ABB92294

DT 31-MAY-2002 (first entry)  
XX

Herbicide active polypeptide SEQ ID NO 1505.

Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.  
YV

PN WO200210210-A2.  
 YY

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.  
VV

PR 28-AUG-2001; 2001WO-EP09892.  
YY

PA (FARB ) BAYER AG.  
VV

Tietjen K, Weidler M;

DR WPI; 2002-269010/31.  
yy

PT Identifying plant target proteins for herbicidally active compounds,  
PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms -

Claim 5; SEQ ID NO 1505; 261pp + Sequence Listing; English

The invention relates to identifying target proteins (AB990790-AB994016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.

SQ Sequence 1164 AA;

Query Match	42.6%	Score 2633.5;	DB 23;	Length 1164;
Best Local Similarity	48.8%;	Pred. NO. 4e-191;		
Matches 568; Conservative	182;	Mismatches 296;	Indels 119;	Gaps 27

105 FLNSHINSGVSGFKCSASLTSLDLSRNSLSGAPVTTLTSLGCSGSLKPLNVSNTLDPFG 164

```

Db 107 YLOGNFS-----SGDSSSSS-----GC----- 124
QY 165 KVGGLKLNLEVLDSANISGANVGVWLSGCGELKHLAISGNKISGDVDSRCVN- 223
Db 125 -----LEVLDSLSNLSLTDSSIVDYVST-CLNLVSVNFSHNKLAKKSSPSAN 174
QY 224 --LEFLDVSSNFSFGIP--FLGDC--SALOHLDISGNKLSGDFSR----- 263
Db 175 KRITTVLSNNRFBDELPFRFIADFPNSLKHLDLSGNVNVGDFSRLSFGLENLTVFSLS 234
QY 264 -----AISTCELKLNLSNQFVGPPIPLP-----LKSLOVLSLAENKFGCEIP 308
Db 235 QNSISGRFPVLSLNCLETLNLSRNSLIGKIPGDDYWGVPQNLRLSLAHNLVSGEIP 294
QY 309 DFLSGACDTLGLDLSGNHFGAVPFGSGCSLLESLSLSSNNSGGLPMDTLKMRGLK 368
Db 295 PELSLCRTLLEVLDSLSNLSLTLGQLPQSGFTSCGSLQSLANLKNKLSGDFLSTVYSKLSRIT 354
QY 369 VLDSFNEFSGELPELSLTLNLSASLTLTDLSSNNSGPIPLPMLCO-NPKNTLOELYLQNG 427
Db 355 NLVLPFNNSISGVPLSLTNGS--NLRVLDLSNNEFTGEVPSGFCSLQSSVLEKLLIANNY 413
QY 428 FTGKIPLPLNSCELSVLSLSPNLSGTPSSLSGLSLKLDLKLMLMBEIPQEL-MY 486
Db 414 LSGTVPELIGKCKSLKTLTDLSPNLTGLIPKEIWTLPKLSLVMMANLTLGTLPESSICVD 473
QY 487 VKTLETLIDFNDLTGEIPSGISNCTNLMWLSLNNRLTGEIPKMGRLLENLAILKLSNN 546
Db 474 GGNLETLILNNLTLGSLPEBSIKTNNLMWLSLSNNLTLGELIPGIGLEKALILQIGNN 533
QY 547 SPSGNIPELGDCLSLMLDLNTLNFCTIPAMFKGSGKLAANFIAGKRVYVTKNDGMK 606
Db 534 SLTGNIPSELGNCKMLIWLDSNNLTLGNLPEGLASQALVWPGSVSGKOPAFVRNEG-G 592
QY 607 KECHGAGNLEFQGRSQRQANLSTRNPNCTISRYGHTSPFPNNSMFMELMSYML 666
Db 593 TDCRAGGIVGVEEGRRAELHEHPMVHSCPKT-RIYSGMTMTMESNCSMTYLDLSYAV 651
QY 667 SGYIPEKISMYLFLINLGNHDSISIPDEVGDLRGHNLIDLSNKLDERIPQAMSLT 726
Db 652 SSSIPGLGYAMGYLOVNLGNHLLTGTIPDSFGKALGVLDLSHNDLQGFPLSGLSLS 711
QY 727 MLTEILDLSNNLSGPIPEMGOFETTPPAFLNNPGLICGYPLRCPDSNADGVAHHQSHG 786
Db 712 FLSDLDVSNNNLTGPPIPCGQULTPPLTRYANNSGLCGVPLPCCSG-----SRPTRSHA 766
QY 787 R-RPDSLASSVAMGLLFFVCJFGLTLVGRERKRKRKEALEMYAEHGNSGDRTYAN 845
Db 767 HPKQSIATGMSAGIVSFMCIMLMALYAPRK-VQKEXKQREXYIBSLPTS-----S 820
QY 846 TNNKLTGVKBAISINLAIEKPLKLTPLADLLOATNGFINDLSLGGGGDYVYKALIDG 905
Db 821 SSMKLSVHEPJSINATIEKPLKLTFAHLLEATNGFSADSMITGGGFGDYVYKALIDG 880
QY 906 SAVAIAKLIVHSGQGRPEPAEMETIGKIKHRLVYLLGYCKYGDRLILVNEFWKGSLE 965
Db 881 SVAIAKLIVTQGGREFAEMETIGKIKHRLVPLLGCKIGERLLVLEYMKGSLE 940
QY 966 DVLOD-PRKGGVKKLSTRKKIAGSARGLAFLHNCSPIIHRDMKSSNVLLDENTLEAR 1024
Db 941 TVLHEKTKKGGIFLIDMSARKKIALIGARGLAFLHSCIPHIHRDMKSSNVLLDDQFVAR 1000
QY 1025 VSDFGMARLMSAMDTLVSSTLAGTGYVPPRYQSGRCSTKDYVSYGVVLELLTGR 1084
Db 1001 VSDFGMARLVSALDTLHLSVSTLAGTGYVPPRYQSGRCSTKDYVSYGVVLELLTSGK 1060
QY 1085 PTDSPDG--NNNLVGVWKO-HAKLRISDVDPPELMKEDPALTEIELQHLKVAVACLDDBA 1142
Db 1061 PIPDEEEDNNLVGMKOLYREKRGAEIIDPELV-TDKSGDVLEHLVYKLSAQCLDDRP 1119
QY 1143 WRPRTVQVWAMPEK-IOAGSGIDS 1166

```

```

Db 1120 FKRPMTIQWMTMFKELQVDTENDS 1144
RESULT 6
ABB91366
ID ABB91366 standard; Protein; 1166 AA.
XX
AC ABB91366;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 577.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidner M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 577; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 1166 AA;
XX
Query Match 42.4%; Score 2623.5; DB 23; Length 1166;
Best local similarity 48.5%; Pred. No. 2.3e-190;
Matches 573; Conservative 190; Mismatches 341; Indels 77; Gaps 29;
QY 33 YRETHQLSKR--DYLPD-KNLLPDW--SSNKNPCPFDDVTCDD-KVTSIDLSKPLNV 86
Db 32 FNEETALLAKRQNSVSKDPNNVLGNMKYRSGSCSWRGVSCDDGRIVGLDIARNGSLG 91
QY 87 GFSAVSSLSLTLGELBELFNSHIN--GSVSGFKCSASLTSIDLSRNSLGEVTLTLTSL 144
Db 92 TLNLV--NLRLPQLQVLYIQGNVYFSSGSDSDC--YLOVLDLSNLSIDYSMDYVF 147
QY 145 GSCGELKFLVNSNTLDFPGKVS-GGLKNSLLEVLDSANISGANVGVWLSGCGELK 203
Db 148 SKCSNLSVSNISNNKL--VGKLGAPESLSQSLTTLVDSYNTLSD-KIPESFISDFPASKL 204
QY 204 HLAISGNKISGDVVSVCVNLLEFLDVSSNNSFSTGIPFLGCSALQHLDISGNLSGD-FS 262
Db 205 YLDLTHNNLSGD-----FSDL-----FGICNLTFFSLSQNNLSGDKFP 244
QY 263 RAISCTEELKLNLSNQFVGPPIPLP-----LPLKSLQVLSLAENKFGCEIPDFLSGACDTL 318
Db 245 ITLPNCKRLETLNLSRNNLAKGIPNGEYWGSPQNLKQSLAHNRLSGEIPPELSLCKTL 304

```

QY 319 TGLDLSGNHFGVAVPFFGSCSLLESLALSSNNNSGELPMDTLKXMGKLVLDLSFNEFS 378  
 DB 305 VILDSGNTFSGELPSQFTACVWLNQNLGNVYLSGDELNTVSKITGITVLYAVANNIS 364  
 QY 379 GELPESLTNLASLTLLTDLSSNNFSGPILPMLCO-NPKATLOEYLONNGFTGKIPPLIS 437  
 DB 365 GSVPISTLNGS-NLRKVLDSNGFTGVPSPGFCISQSPULEKILLIANVYLSGTPMELG 423  
 QY 438 NCSELVSLHSPNYLSGTPISLSGLSKLDKLMNLMLGEIPELMTYK-TLETLIL 495  
 DB 424 KCKSLKTLIDLSFNNELTGPIPKIEMWLPISDLVMMANNLTGTIE-EGCVKXGMLLETLIL 482  
 QY 496 DFNDLTGEPISGLSNCCTMLNWSISNNRLTGEIPKATGRLENLAILKLSNNSFGNIPDE 555  
 DB 483 NNNLLTSGIPISIRCTNMWISLSNNRLTGKIPSGISGKSLAILLOLGNNSLSGNVPRQ 542  
 QY 556 LGDRSLIWLDTNLNLFNGTIPAMFGKSGKIANFLAGKRYVYIKNDGKKECHGAGNL 615  
 DB 543 LGNCKSLIWLDTNLNLLGDLPGELASQAGLVMPGVSQKQFAVNEG-GTDRGAGGL 601  
 QY 616 LEFGIASEQNLRLSTNPNCTISRVYCGHTSPTEDNNGSMFLDMSYNNLSGYIPEKIG 675  
 DB 602 VEEFGIASERLERLPMVHSCPAT-RIYSGMTMYFSANGSMIYFDISYNAVSGPIPGYG 660  
 QY 676 SMPFLINLNGNDISGSIPEVGLRGLNMLDLSNKLDRIPQMSALTMLEIDLIN 735  
 DB 661 NMGLQVNLNHRNRTITIDSPGSLAIGVLDLSHNNLQGYLPGSLISLSDIDVSN 720  
 QY 736 NNLSGPIPEMGQETEPFAFLNPNGLCGVPLPCDPSNADGVAHHQSRGRPASIASG 795  
 DB 721 NNLGPIPEFGQULTFPVSRVYANNISGLCGVPLRPC-GSAPRRPISRIHAKK-QVATA 777  
 QY 796 VANGILFSPFCITGLIVGEMKRRKKEAELEMAEAGNSGDRANTNNTKLTGYKE 855  
 DB 778 VIAGIABSPFCVWMLVALYVRK-VQKKEQKREKYEISLPTSG-----SCSWKLSVPE 831  
 QY 856 ALSINLAPEKPRKLTFLADLQATNGFNHDSLISGSGFDVYKALIKGSAVAIKLH 915  
 DB 832 PLGINVATPEKPRKLTFLADLQATNGFNHDSLISGSGFDVYKALIKGSAVAIKLH 891  
 QY 916 VSGQGRPEFAENETIGIKIHRNLVPLLYGCKVYDERLLVNEVMKYSLEVDLQD--PRK 973  
 DB 892 ITQGRPEFAENETIGIKIHRNLVPLLYGCKVYDERLLVNEVMKYSLEVDLQD--PRK 951  
 QY 974 GGKYLKLSRTRKIAISANGLAFLHNCSPHITHRMKSNVLLDENLEARVSDFGMARL 1033  
 DB 952 GGKYLKLSRTRKIAISANGLAFLHNCSPHITHRMKSNVLLDENLEARVSDFGMARL 1011  
 QY 1034 MSAMDTLHLSVSTLACTPGVVPPEYQSPFCSTKGVYSGVAVLLELTGKRPDSDPG- 1092  
 DB 1012 VSAIDTLHLSVSTLACTPGVVPPEYQSPFCSTKGVYSGVAVLLELTGKRPDSDPG- 1071  
 QY 1093 DNNLVWVWYQ-HAKLISVFPDELMKEDPALEIELLQKVAVALCLDDRAMRPTMYOV 1151  
 DB 1072 DNNLVWVWYQ-HAKLISVFPDELMKEDPALEIELLQKVAVALCLDDRAMRPTMYOV 1130  
 QY 1152 MAMFEKIQAGSGIDSGSTIRSIJEDGFSITENWMSIXEVP 1192  
 DB 1131 MAMFEKEMKADTEDE-----SLDEPSLXETP 1156

RESULT 7  
 ABB91692  
 ID ABB91692 standard; Protein; 1143 AA.

AC ABB91692;  
 XX  
 DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 903.  
 XX  
 KW Herbicidal; plant; agriculture; herbicide.

XX  
 OS Arabidopsis thaliana.  
 XX  
 PN MO200210210-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PR 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PA (FARB) BAYER AG.  
 XX  
 PI Tiejfen K, Weidner M;  
 XX  
 DR WPI, 2002-269010/31.  
 XX  
 PT Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -  
 PS  
 PS Claim 5; SEQ ID NO 903; 261pp + Sequence listing; English.  
 XX  
 CC The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.  
 CC  
 CC Sequence 1143 AA;  
 SQ  
 Query Match 40.6%; Score 2509.5; DB 23; Length 1143;  
 Best Local Similarity 46.2%; Pred. No. 1.1e-181;  
 Matches 556; Conservative 183; Mismatches 353; Indels 111; Gaps 25;  
 QY 9 LSVTLTLPFFSPFSLSPQASPSQSIYREIHQISFKVLPD-KNLDPMSNKNPCTFDG 66  
 DB 14 IQISFIFLTLHLSQS-SSSQSSSLKTLSTLSLRFKWTIQQDPNNILNWSPPKPCQFSG 72  
 QY 67 VTCRDKVTSIDLSKPELN--VGFSAVS---SLSLTGLSPLFNSHINSVSGFKC 120  
 DB 73 VTCRDKVTSIDLSKPELN--VGFSAVS---SLSLTGLSPLFNSHINSVSGFKC 125  
 QY 121 SASLTSIDLSRNSISGVTTLTSLGSCSGKPLNVSNTL--DPRKYSGGKILNS--L 175  
 DB 126 PLTTLHLELSSGLIG-----TLPENFESKYNLISITLVSNNFTGKLPNDLFLSSKCL 179  
 QY 176 EVDLSANSISGANVWVWVLSGCGELKHLAISGNKISGVVDSRCVNEFLDVSSNNFS 235  
 DB 180 QTLDSLNNITG-----PISGLTIP----- 199  
 QY 236 TGPIFLDCCALQHLIDISGNKLSGDFSPRAISTCTEKLINISNQVGPILPL--PLKSL 293  
 DB 200 -----LSSCVSMYTYLDRSGNSISGYISDLINCTNLSKSLNLSNNNDGQIPKSRGLKLI 254  
 QY 294 QYLSLAKNTGTGPIPLPSGACDTLTGLDLSGNHFGVAVPFFGSCSLLESLALSSNNFS 353  
 DB 255 QSLDLSNNRLTGMIPPEIGTCSRLONLSTYNNFTGVPISLSSCSMWQSLDLSNNNIS 314  
 QY 354 GELPMDTLKMRGLKLVLSFNEFSGELPESLTNLSA--SLTLIDLSNNSPGLIPNLIC 411  
 DB 315 GPPPNLTILSFGSLQILLSSNNLISGDFP---TSISACSLNIADPSSSRFSGVLPDLC 371  
 QY 412 QNPKNTLOEYLONNGFTGKIPPLISNCELSVLSLSPYLSGTPISLSGLSKLDKLMNL 471  
 DB 372 PGAA-SLEELRLPDNLVTEIPPAISQCELELTIDLSLNYLNGTIPPEIGNQKXEGFIA 430  
 QY 472 WLNMLEGEIPQELMTYKLTETLIDPNDLTGEPISGLSNCCTMLNWSISNNRLTGEIPRW 531



QY 481 POELMYVKTLETLIDPNDLTGEIPSGISNCTUNLMSISNNRLTGEIPKIGLENLAI 540  
 Db 434 PPEIGKCKNLKLLLNHNHKGELPTELFCNSNLEWISLTSNELTGEVREVLRLAV 493  
 QY 541 LKLSNNSPSGIPDELDCRSLIWLDTNLTFNGTIPAAEFKSG-KIANPFIAGRYV 599  
 Db 494 LQLNNSISGQIPWELNCSIMLDSSNKLGEIPRLGRQIGAKPGIPSGTIVF 553  
 QY 600 IKNDGKKECHGAGNLLFQGISBQNLRLSTNPNCTISRVYGHTSPTFNNGSMFL 659  
 Db 554 VRNVG--NTCKGVGGLLEFAGIRPERLLQVPSLRTCNF-ARMYSGPILSKFTTYQVEL 610  
 QY 660 DMSYNMLSGYIPKESIMPLFIANGNDISGIPDEYVDLFGNLTDLSSNKLDGRIP 719  
 Db 611 DLSNDRGKIPPEFEGDVALQVLELSYNQISGIPSSIGRLKDLGVFASHNRLLQGIIP 670  
 QY 720 QMSALTMLTEIDLSNNNLSGPIPEMGQFETPPAPFLNPPGCGVPLRRC-----DP 772  
 Db 671 DSFSNLSFLVQIDLSYNDLTGQIPQSGQLSTLPASQYEHNPGLGVPLPECRNNEPETNP 730  
 QY 773 SNADGVANHQSHGR-RP--ASLAGSVANGLLFSFVCFGLILVGRNKRKRKKEAELE 829  
 Db 731 DVAMG-----TGRAKPGTASWANSIVIGILISIASICILTWGIAHSRKEAF-EMK 783  
 QY 830 MYAEGHNSGDPRTANNNTMKTGVEKALSTINLAPEKPKLTPADLIQATYFNDSLI 889  
 Db 784 ML-----KSIQAAHATTWKIDKKEKPLSINAVTPQRKLKLFSQLLEATNGFSABSLI 838  
 QY 890 GSGGFGDVAIKKDSNAVAIKKLHVSGQDREFMAEMETIGIKHNVLVPLGYCYG 949  
 Db 839 GCGGFSEVFAKTKDSNVAIKKLIRLSYQGRREFMAEMETIGIKHNVLVPLGYCYG 898  
 QY 950 DEBLVNEWKVKSIEDVQDPKGGVK-----LKLSTRKIAIGARGLAFLHNCSP 1003  
 Db 899 KERLLVYEFERGESLEEMH---GRVKTQDRRIITWERKKIARCAKGCFLHNCIP 954  
 QY 1004 HIHRDMKSNVLDENLEARYSDFGMARLMSAMDTLVSSTLACTPGVPEYVQSRIC 1063  
 Db 955 HIHRDMKSNVLDHMDARVSDFGMARLISALDTHLSVSTLACTPGVPEYVQSRIC 1014  
 QY 1064 STYGDVYSIGVVLLELTGKRPTSPDFGDNVLVGVYQKAKL-RISDVPPELM----- 1117  
 Db 1015 YAKGDVYSIGVVLLELTGKRPTSPDFGDNVLVGVYQKAKL-RISDVPPELM----- 1117  
 QY 1118 KEDPALEI-----ELLQHLKAVACLDDBARRRPTMVAAMFEIAGS-GIDSOST 1169  
 Db 1075 QERARAEAAEVVKEWRYMETITLQCVDPFSRRPTMLQVYAMRELVPFGSTVSSNST 1133  
 K 9  
 AAB25323  
 ID AAB25323 standard; Protein: 919 AA.  
 AC AAB25323;  
 XX 27-NOV-2000 (first entry)  
 DT  
 DE Eucalyptus grandis cell signalling involved protein SEQ ID NO:642.  
 XX  
 KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
 KW plant cell signalling; modulation; transgenic plant; pathogen; growth;  
 KW environmental change; development; cell proliferation; differentiation;  
 KW elongation; survival; disease resistance; nutrient metabolism.  
 OS Eucalyptus grandis.  
 XX  
 PN MO200042171-A1.  
 XX 20-JUL-2000.  
 PD 11-JAN-2000; 2000MO-US00724.  
 XX 12-JAN-1999; 99US-0228986.  
 PR

PR 01-NOV-1999; 99US-0162866.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 XX  
 PI Strabala TV, Nieuwenhuizen NJ;  
 XX  
 DR WPI; 2000-476052/41.  
 PT Isolated polynucleotide encoding a polypeptide involved in cell  
 PT signaling used for generating transgenic plants with modified responses  
 PT to external signals -  
 XX  
 XX  
 PS Claim 3; Page 291-293; 527pp; English.  
 CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
 CC are involved in cell signalling. The polynucleotide and protein  
 CC sequences can be used to modify the response of plant cells to external  
 CC signals e.g. environmental changes or pathogens during the growth and  
 CC development of a plant. They can be used to modify cell proliferation,  
 CC differentiation, elongation and survival, resistance to disease and  
 CC nutrient metabolism. Examples of modifications which can be produced are  
 CC altered fruit ripening and senescence of leaves and flowers e.g. to  
 CC delay senescence and prolong the life of cut flowers or enhance  
 CC modifications of reproductive organs to engineer sterile plants. Other  
 CC organs providing fruit and vegetables which have a longer shelf life  
 CC between harvest and consumption, or to decrease branching frequency in  
 CC forest tree species giving long stretches of valuable knot-free clear  
 CC wood which can be used in solid timber furniture and veneers.  
 XX  
 SQ Sequence 919 AA;  
 Query Match 35.4%; Score 2189.5; DB 21; Length 919;  
 Best Local Similarity 50.6%; Pred. No. 1.7e-157;  
 Matches 475; Conservative 134; Mismatches 276; Indels 53; Gaps 18;  
 QY 264 AISTCETELKLNINSGVGPVPL--PLKSLQVLSLAENKFTGEIPDLGACDTLTGL 321  
 Db 3 SFSNCTSLHTSLANNFTGEIPHSFRLASLQKLDISHNHLTGWIPDELGTACNSLIEL 62  
 QY 322 DLSGHHFYGAVPPFGSCSLLESIALSSNNFSGELPMDTLTKRGLKVLDSNERSGEL 381  
 Db 63 TLFNNISGPIPLFSSGSMWQSIDLSTNNISGEPDSIRAKLSLESLSLNNNTISGLF 122  
 QY 382 PESLTNLSASLTLDSSNFSGPILNLCQNPKNLTQELYLQNGEFTGKIPPLSNCSB 441  
 Db 123 PASL-SYCKRLRIIDFSSNCFSGIIPDLCLGAA-LLEELRAPDNLITGNIPQLSQCSQ 180  
 QY 442 LVSLHLSFNYLSGTIPSSLSGLSLRDLKLMNLGELIPQELMYVKTLETLIDPNDLT 501  
 Db 181 LKTIIDPSLNTLVGSIIPAEIGLNTLEQLIMFNGLGEIPPEIGKCKNLKLLNHNHKL 240  
 QY 502 GEIPSGLSNCTNLNWLISLNNRLTGEIPKWIIGLENLAILKLSNNSPSGIPDELDCRS 561  
 Db 241 GEIPPELFNCSNLEWISLTSNELTGEVREVLRLAVLQGNNSLSQIPWELNCSG 300  
 QY 562 LIMPDLMTNLFNGTIPAAEFKSG-KIANPFIAGRYVYIKNDGKKECHGAGNLLFQGS 620  
 Db 301 LMTDLSSNKLGTGIPRLGRQIGAKPGIPSGNTLVVRNVG--NTCKGVGGLLEFAG 358  
 QY 621 IRSBQNLRLSTNPNCTISRVYGHTSPTFNNGSMFLDMSYNMLSGYIPKESIMPL 680  
 Db 359 IRPRLQVPSLRTCNF-ARMYSGPILSKFTTYQVELDSDNQLRGKIPPEFEGDVAL 417  
 QY 681 FIINLGHNDISGIPDEYVDLFGNLTDLSSNKLDGRIPQMSALTMLTEIDLSNNNL 740  
 Db 418 QVLELSTNQLSGEIPSSIGRLKDLGVFEASHNRLLQGIIPDSFSNLSFLVQIDLSYNDLTG 477  
 QY 741 PIPMGQFETPPAPFLNPPGCGVPLRRC-----DPSNADGVANHQSHGR-RP--A 790  
 Db 478 QIPQGLSTLPASQYEHNPGLGVPLPECRNNEPETNPDVAMG-----TGRAKPGT 531

QY 791 SLAGVAMGILFVFCILFGLILVGRMRKRRKKEALEMVAEAGHNSGSDRTANNIMKL 850  
 Db 532 SWANSIVLGLISIASICILTVGIMRSRRKAE-EMKML-----KIQAHATTWKI 585  
 QY 851 TGVEALISIMAAFEKPLKLTPEADLLQATNGPHNDLSIGSGFDVYKAILKDSAVAI 910  
 Db 586 DKREPELSIVAFQRLRLKFSQLEATNGFSASLIGCGGFGVFAATLKDGSNVAI 645  
 QY 911 KXLIHSGGDRPMAMETITGKIKRNLVPLGYCKVGBERLLVEMKYSLEPVLGD 970  
 Db 646 KXLIHSGGDRPMAMETITGKIKRNLVPLGYCKVGBERLLVEMKYSLEPVLGD 704  
 QY 971 PKGGVY-----LKLSTRKIAIGSARGLAFLHNSCPHIIHRDKMSNVLLDENLEAR 1024  
 Db 705 ---GRVITQDRRLITWEERKKIARGAKGICPLHNCIPHIIHRDKMSNVLLDHDMEAR 761  
 QY 1025 VSDGFMARLMSAMDTLHLSVSTLACTPGVPEYIYQSFRCTKGVYISYGVILLELTGR 1084  
 Db 762 VSDGFMARLISALDTHLSVSTLACTPGVPEYIYQSFRCTKGVYISYGVILLELTGR 821  
 QY 1085 PTSDPDGDNNTLVGMVGMHKL-RISDVPEPELM-----KEDPALER-----ELQHLK 1132  
 Db 822 PTDKEDFDGDTLVGMVGMHKLREGKQMEYIDPAMSTKVGEBEAEAEAVEMRYME 881  
 QY 1133 VAVACLDDBRAMRPTMVOVMAPKEIQAGS-GIDSOST 1169  
 Db 882 ITLQCVDDPSPRRPTMLQVAMLELVPSTVSSNST 919

RESULT 10  
 ABB93388  
 ID ABB93388 standard; Protein; 1192 AA.

AC ABB93388;  
 XX 31-MAY-2002 (first entry)  
 XX 31-MAY-2002 (first entry)  
 DE Herbicidally active polypeptide SEQ ID NO 2599.  
 XX Herbicidally active polypeptide SEQ ID NO 2599.  
 KW Herbicidal; plant; agriculture; herbicide.  
 OS Arabidopsis thaliana.  
 XX MO200210210-A2.  
 XX 07-FEB-2002.  
 XX 28-AUG-2001; 2001WO-BP09892.  
 XX 28-AUG-2001; 2001WO-BP09892.  
 PR 28-AUG-2001; 2001WO-BP09892.  
 XX (FARB ) BAYER AG.  
 XX Tietjen K, Weidner M;  
 XX WPI; 2002-269010/31.  
 DR Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -  
 XX Claim 5; SEQ ID NO 2599; 261pp + Sequence Listing; English.  
 XX The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.  
 XX Sequence 1192 AA;  
 SQ  
 Query Match 26.0%; Score 1608.5; DB 23; Length 1192;  
 Best Local Similarity 35.2%; Pred. No. 4.4e-113;  
 Matches 453; Conservative 177; Mismatches 421; Indels 235; Gaps 36;  
 QY 11 VTTLFFSFSLSQASPSQSLREIHOQLSPFDVLPDKKLLDWM--SSNKNCTFDGVT 68  
 Db 4 LTLALFLFF--SPSSAIVDLSETTSLSPKSLNPELSLSSMNVSSASHCWDVGT 61  
 QY 69 CRDKMYSIDLSKPLWGFSAVSSSLSTGLSEPLSNHNGSVS---GFKCSASL 124  
 Db 62 CLIRGVNSLSPSLSLK---CQIPKEISLKLRELCLAGNQSFGKIPPEIMWLK--HL 115  
 QY 125 TSLDSRNSLSGPVTTTSLSCSGGLKFLVNSNTLDFPKV--SGGLKLSLEVLDLSA 182  
 Db 116 QTLDSGNSLTGLLPRLL--LEELPOLLYLDLSDN--HFSGLPPSPFTSLPALSSLDVSN 171  
 QY 183 NSISGANVGVVLDDGGELKHLA--ISGNKISGV----- 216  
 Db 172 NSLSGE-----IPPEIGKLSNLNVLGNLNSFGQIPSEIGNISLKNFAPECFENG 225  
 QY 217 ---DVSRCVNIFFLDVSSNNPSTGI-----PFLGDCSALQ 248  
 Db 226 LPKEISKXKLAKLDLSYNPCKSIPKSGFGLHNLISLIVASBELIGLIPBELNCCKSLK 285  
 QY 249 HLDISGNKLSGDFSPRAISTCTELKILNLS--NQFVGPIPL--PLKSLQVSLAENKFT 304  
 Db 286 SIMLSFNSLSGFLPLELS--EIPLLTFSAERNQSLSPSMGKMKYLDLILANNRFS 342  
 QY 305 GRIP-----DPLSGA-----CDTLTGDLPSGNHFFYGAVPPEFGSCSL 341  
 Db 343 GELPHEIEDCPMLKHLASNLSSGIPRELQSGSLAIDLQSLTSGTIEEVEDGSS 402  
 QY 342 LESLASSNNSFSGELPMDTLKMRGLKVLDSFNEFSELPESL--TNL----- 388  
 Db 403 LGEILLITNNQINGSIPED-LMKL-PLMALDLSNNFTGEIPKSLTNLWELFASVRL 460  
 QY 389 ---SASILLTDLSSNNSFSGILPMLCQPKNTLOELYLQNNGFTGKIPPTLSN 438  
 Db 461 EGYLPKEIGNAASLKLVLSDNQLTGEIPREI--GKLSLSTVLNANMFPQKIPVELGD 518  
 QY 439 CSELVSLHSFNYLSGTTIPSSLSLSKLDLKLWMLGEBIPQ-----ELMY 486  
 Db 519 CTSLTTLDLGSNNLQGIQDXTALAQCLVSYNNLSGSIKPSKSAFYHOIEMPDLSF 578  
 QY 487 VKTLETLILDENDLGEIPIGSLNCTNINWISLNNRLTGEIPKXIGRENLATILKSN 546  
 Db 579 LQHHGIFDLSYVRLSGPIPEELGECVLVEISLNNHLSGELPASISRLTNLTIIDLSGN 638  
 QY 547 SPSGNIPDELGDCRSLIMDLNTNLFNGTTPAMFKOSGKIAPNFAKRYVYIRKNGK 606  
 Db 639 ALTGSIPKEMGSLKQGLNLIANNQNGHIP----- 669  
 QY 607 KECHGANLLEFOGIRSEQLNRLSTRNPNITTSRYVGGTSPFPDNGSMFLDMSYML 666  
 Db 670 -----ESPGLLSLVKLTNTKMKLDGVPAISGLKELTHNDLSFNNL 712  
 QY 667 SGYIPKATIGMPFLFLNLGNHNSIGSIDPEVGLAGNLTIDLSKGLGRIPQAMSAIT 726  
 Db 713 SGEISSELSTMEKLVGLYIRKQFTGBISELGNLQLEYLDVSENLGELIPTKICGLP 772  
 QY 727 MLTEIDSSNNLSGPIPEMQFETFPAPKFLNPNLGCYPLPFCDDPSNADGAHQRSHG 786  
 Db 773 NLEFLINAKNNLGEVPSDVCQPSKALLSGNKEICG---RVVGSDC-----KIBG 821  
 QY 787 RRPASLAGVAMGILFS--FVCIPLGL--ILVGRMRK---RRKKEALEMVAEAG--HG 836  
 Db 822 TKLASAGIAGLMLGFTIIVFVFSLRRAWMTKRYKQRDPERMEBSRLKGVVDQNLV 881  
 QY 837 NSGRTANNNTWKLTGVKALISIMAAFEKPLKLTPEADLLQATNGPHNDLSIGSGFGD 896



Db 882 LSGSS-----REPISINAMEEOPPLKVRIGDIVEADHFSKNIIGDGGFGR 930  
 Qy 897 VYKAILKDGSAVAIKLIVHSGOGDREFMAEMETIGIKRNLVPLLYGKVDRELLN 956  
 Db 931 VYKALPEKTVAVAKLSEAKTQGNREFMAEMETIGIKRNLVPLLYGKVDRELLN 990  
 Qy 957 EVMYKGSLEDTLOPKKGVKALKSTRKLAIGSARGLAFHNCSPHIIHRDKSSNVL 1016  
 Db 991 EVMYKGSLEDTLOPKKGVKALKSTRKLAIGSARGLAFHNCSPHIIHRDKSSNVL 1050  
 Qy 1017 LDENLEAVSDPFGMARLMSAMDTHLSVSTAGTPGVYPEYVGSFCCKKGVYSGVYL 1076  
 Db 1051 LDGFEPRVADFGARLISACEHSV-TVLAGTFTGYPPEYVGSFCCKKGVYSGVYL 1109  
 Qy 1077 LELLTGKRPDSDPFGN--NIVGWYKQAKL-RISDVPPELMKEDALETELQHLK 1132  
 Db 1110 LELVTGKEPT-GDPFKSEEGNLVGMALIKINGQKAVDVIDPLV--SVALKNSQRLIQ 1166  
 Qy 1133 VAVACLDDBAMRRPTMOWAMPEKEI 1158  
 Db 1167 IAMCLAEETAKRPMMLDLKALKEI 1192

RESULT 11  
 ABB93804  
 ID ABB93804 standard; protein; 1036 AA.

AC ABB93804;  
 DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 3015.

KM Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

XX Tietjen K, Weidner M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,  
 comparing aligning and comparing nucleic acid or amino acid sequences  
 from plant with nucleic acid or amino acid sequences from non-plant  
 organisms -

PS Claim 5; SEQ ID NO 3015; 261bp + Sequence Listing; English.

CC The invention relates to identifying target proteins  
 (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 aligning and comparing nucleic acid or amino acid sequences from plant  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.

XX Sequence 1036 AA;

Query Match 22.7%; Score 1402.5; DB 23; Length 1036;  
 Best Local Similarity 33.4%; Pred. No. 1.7e-97;

Matches 394; Conservative 182; Mismatches 416; Indels 189; Gaps 33;  
 Qy 11 VTTLPFFSPFSLFQASPSQSLVREIHOLISFDVLPDQNLDPWSSNNPCTPGVTCR 70  
 Db 4 ILLVFPVGSVSEOPCHPN-----DLALRELAGALKNKSVESWLNRCCEMVGCE 58  
 Qy 71 ----DDVTSIDLSKRPVAVGFVAVSSSLITGLESLFLSNHNGSVGPKCASLTS 126  
 Db 59 GSDVSGVTVTLVPEKGLR--GVISKSLGELT-----ELRV 92  
 Qy 127 LDRSRNLSGPPVTTLTLSGSCGLKFEINVSNTLPFGKVSGLKLNLEVLDSANST 186  
 Db 93 LDRSRNLSGPPVTTLTLSGSCGLKFEINVSNTLPFGKVSGLKLNLEVLDSANST 125  
 Qy 187 GANVVGVLSDGCGELKHLAISGNTSGDV-DVSRVNLLEFLDVSNNPSTGI-PLGDC 244  
 Db 126 G-SVLGV--SGKLQSLNLSNLSGKLSDVGFVGLVNLVNSNNLFEGBIHEL--C 180  
 Qy 245 SA---LQHLDSGNKLSGDFSRATSTCELKLNLSNNOFVGPFP--LPLKSLQYSLA 299  
 Db 181 SSSGGIQLVDLSNRLVGNLDGLYNCSKSIQQLHDSNRLTGQPLDYYSIRELSLS 240  
 Qy 300 ENKFTSEIPDFLSGACDTLTGLD--LSGNHFGAVPFPFGSCSLTESLSSNFGEL 356  
 Db 241 GNYLSGELSKNLS---NLGKLSLISENRPDVTLPVFNLTQLEHLDVSSNKFSGRF 296  
 Qy 357 PMDTLKRMGLVLDLSNRFSGELPESLTLNLSASLTLDLSNNPFGPIIPNLCNPKN 416  
 Db 297 P-PSLQCSKLVLDLRNNSLSGSLNINLFTGFT-DVLCLDLSNHFSGPLPDSIGHCPK 353  
 Qy 417 TLQELVQNGFTGKIPTLSNCSLVSILHSFVNLGSLTSSLSGLSKLRLKMLNML 476  
 Db 354 -WKILSLANFRGKIPTLPKQLSLFLSLSN-----SFVDSSTNVL 398  
 Qy 477 EGBIPELMYVTTLTLLDPRDLTGEIPSGLSNCTNMNLSNNLJTEBIPKIGRL 536  
 Db 399 Q-----HCRNLSTLLSKNFGEBEPNNVTGFDNLATLALNCGLRQIPLMNLCK 450  
 Qy 537 NLAILKLSNNSPSGNIPELDCRLIWLIDLNTLFPNGTIPAMFKSGKIANFLAGKR 596  
 Db 451 KLEVLDSNHFYGTIPWIKMESLFYIDPSNTLTGALPVAITELKNILRUNGASQM 510  
 Qy 597 Y-----VYIKNDGKKECHAGNMLEFGIIRSEQLNRLSTRPCNTSR---VVGHT 646  
 Db 511 TDSGIPLYVR-----NKSNGLPYNOVSFPFSIT--- 542  
 Qy 647 SPTFDNNSMFLDMSTYMLSGYIPKATISNPIYFIINLGHNDISGIPDEVDLRLINI 706  
 Db 543 -----LNNRNLNGTILPEIGRLKEIHLMLDLSRNNFTGTIPDSIGLNLLEV 588  
 Qy 707 LDLSNKLDRIPQAMSLTWTLEIDLSNNLSGPIPEMGOFEFPFAKFLANNPGLCGTP 766  
 Db 589 LDLSNKLDRIPQAMSLTWTLEIDLSNNLSGPIPEMGOFEFPFAKFLANNPGLCGTP 766  
 Qy 767 LPRCD-----PSNADGYAHQSHGRRPASLGSVAMGLIFSVICFGLILVGERKKR 821  
 Db 649 DSPCDVLSNNLNFKSGSRNNNGKFGKRSIYVLTSLAIGTILLVLTLL-----RI 702  
 Qy 822 RKKAELMVAEAGHNGSGDRTANNMTKLTVGEALS-INLAFEK-PLRKTFLADLQA 879  
 Db 703 SRKVD-----DRINDVDETTISGVSKALGPSEKIVLFHSCGCKDLSVELLS 750  
 Qy 880 TNGHNDLSIGSGFGDGYKALIKDGSAAVAIKLIVHSGOGDREFMAEMETIGIKRNL 939  
 Db 751 TNNFSQANITIGCGGFLGYKANFPDGSKAAVRLSDCGQMRERQAEVRLSRHKNL 810  
 Qy 940 VPLIGYKVDDELLVNVWVKYSLSDVYQDPKGGVYKALKSTRKLAIGSARGLAFH 999  
 Db 811 VSLGCKHNDRLIYSFEMENGLDYLWHERVDGMMTLIMDVRIKIQAGARGLAYLHK 870  
 Qy 1000 NCSPHIIHRDKSSNVLDENLLEAVSDPFGMARLMSAMDTHLSVSTAGTPGVPEYQ 1059  
 Db 871 VCEPNVIHRDVKSNLILDEKPEAHLDLRLRPYDTHVT-TDLVGTIGVLPPEYSQ 929

QY 1060 SFRCSSTKGVSYGVVLLLELLTGKPTD-SDFGDNMLVGMWKO-HAKLRISDFPELM 1117  
 DB 930 SLIATCRGVSYGVVLLLELVGRRPEVCKGKSCDLVSRFQMKAEKREMLDITI- 988  
 QY 1118 KEDPALEIHLLOHLKVAVACLDPRAMRPTWQVYMAFKEI 1158  
 DB 989 -RENNVRETVLEMLEIACKCIDHEPRRRPLIEBVTWMLDL 1028

## RESULT 12

ABB93939  
 ID ABB93939 standard; Protein; 1102 AA.

AC ABB93939;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 3150.

Herbicide; plant; agriculture; herbicide.

Arabidopsis thaliana.

MO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB) BAYER AG.

Tietjen K, Weidner M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicidally active compounds,  
 comprising aligning and comparing nucleic acid or amino acid sequences  
 from plant with nucleic acid or amino acid sequences from non-plant  
 organisms -

Claim 5; SEQ ID NO 3150; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins  
 (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 aligning and comparing nucleic acid or amino acid sequences from plant  
 with nucleic acid or amino acid sequences from non-plant organisms using  
 suitable search parameters, where plant sequences having an E-value  
 greater by a factor of 3 than the E-value of most similar non-plant  
 sequences are selected. The polypeptides or nucleic acids encoding them  
 are useful for identifying modulators. The identified modulators are  
 useful as herbicides.

Sequence 1102 AA;

Query Match 22.3%; Score 1380.5; DB 23; Length 1102;

Best Local Similarity 32.0%; Pred. No. 8.8e-96;

Matches 387; Conservative 179; Mismatches 470; Indels 173; Gaps 31;

QY 1 MKTFFSFPISVTTLFFSFPFSLFQASPSQSLYREIHOLISFDVLPD-KNLLPDMSSNK 59

DB 5 MMLAVFFISLILILI-----SETTGALBEGQYLLEIKSKFVDKAKONLRMNSND 55

QY 60 N-ECTPGVTC---RDDKVTSLDSSKPLNVGFSVSSSLSLTGLIESLFLNSHINGS 114

DB 56 SVFCGRTGWCNSVSDPEVTLNLSMSVL-----SGRLSPS 92

QY 115 VSGFKCSASLTSIDLSRNSLSGVTTLTSLGSCGLKFLNVSNTLDPFGKYSGLKINS 174

DB 93 IGGI---VHLKQDLISYGLSGKIP--KEIGNCSSLEILLKLNNOFDEIPEIG-KLVS 146

QY 175 LEVLDLSANSISG-----ANVYGMVLSDGGELEKHLA---ISGNIS 213

DB 147 LEMILYNNRISGSLPEVEIGNLISLQLYTYSNNISGQIPRIGNLKLITSPRAQNMIS 206

QY 214 GDV--DVSRCVNLEFDVSSNNFSTGP-FLGDCSALOHLDISGNLSDFSRAISTCTE 270

DB 207 GSLPSEIGGESILVLMGLAQNQSLGELPKRIGMLKKSQVILMENEFSGFIPREISNCTS 266

QY 271 LKLNISNQGFGVGPPL--PLKSLOTSLAENKFTGEIPDLGACDITLGLDLSGNHF 328

DB 267 LETLALYKNOLVGPVPELQDLSLEFLYLRNGLNGTIPREI-GMLSTAIEDFSENNL 325

QY 329 YGAVPPFGSGSLLESALSSNNPSGELPMDTLLKMRGLKVLDSFENRSGELPESLITNL 388

DB 326 TGEIPLEIGNEGLELLYLPENQLTGTIPE-LSLTKNLKSLDLSINALTGPIPLGFOYL 384

QY 389 SASLITLDSNNFSGPILPNLQNPNTLQELYLQNGFTGKIPLPLSCSLVLSLHS 448

DB 385 -RGLFMLQLFQNSLSGTIPKTL--GWSYDLVLDMSDNHLSGRIPSYLCLHSNNITLNLG 441

QY 449 FNYLSGTIPSSLSGLSKLDKLMNLBGEIPELWYVKTLETLIDFNDLTGEIPLSG 508

DB 442 TNNLSGNIPTGITTCTIVQRLKRNVLVGRFSPNLCKQVNVTAIEIGQRPFGSIPREV 501

QY 509 SNCTNLNWSISNNRNGEIPKWIIGLENTALILKLSNNSFGSNIPELGDRCSLIWLIN 568

DB 502 GNCSALQRLQDLNCGFQELPREIGMLSQGLTNISNKLGEVSEIPNCKMLQRLDMC 561

QY 569 TNLFNGTIPAMFKOSGKIAPNFTAGKRYVYIKNDGKKCKGAGNILEFGGISBOLNR 628

DB 562 CNPFSGTLPSEV-----GSLYOL-----BLK 583

QY 629 ISTRNPNITSRVVGHNSTPTFDNNGSMWFLDMSYNNLSGYIPKEISMPYLI-LNLGH 687

DB 584 LSNNN-----LSGTIPVALGNLSRLTELQNGNLENGSIPRELSGLTGLQALNLISY 635

QY 688 NDISGSLPDEVGLRGLNILDSSNKLDRIPQAMSAITMLTEIDLSNNLSGPIPEMGQ 747

DB 636 NKLTLIPPELSNMLVLEFLLNNNNLGSEIPSSFANLSLIGVFSNLSLGTPL-- 693

QY 748 FETPPAKLNNPGLCYVPLPRC-----DPSNADGVAHQSRHGRPASLGSVAMGL 801

DB 694 -RNISMSFIGNELCGPLNQCITOTOPAPSQITG-----KGGMRSSKIIAII 742

QY 802 FSPVCIFFILVGERKKRRKKEAELEMYAEGHNSGDRTANTMTKLTVYKALSLNL 861

DB 743 AAVTGVSLMLTALYYLMRRPVRTVASAQDG-----QPEMSL 782

QY 862 AAFKPLRLKLPADLIQATNGRHNSLISGSGFDVYKAILKDSAAVAIKKLI--HVSQ 919

DB 783 DIYPPKEGTFQDILVAAITDNDPESFVVRGACGVYKAVLPAGTTLAVKKLASHNHEGN 842

QY 920 G--DEEFAEMETIKIGHNLVPLLYGCKYGDERTLIVNEVMKYGSLEVDLQDPKGV 976

DB 843 NNNVDSFRAEITLTGNIRKNIYKLFHRCNNHGSNLLIYEMPMKSLGELIHPD---SC 899

QY 977 KKLSTRRKLIAGARGLAFLHNHCSPHIIRDMKSNVLLDENLFAKVSDDGMALMSA 1036

DB 900 NLDMSKRFKIALGAAGLAIHDDCKPRIFHRDIKSNLILDDKFAHAGDGLAVYD- 958

QY 1037 MDTHLSVSTIAGTPGVYPPYYSFPCSTKGVYGVVLELLTTSKRPDSDPFQDNML 1096

DB 959 MPEKSMASALAGSYIAPYATMKVTEKSDYISGVVLELLTTSKAVQPIPDQ--GVY 1017

QY 1097 VGVYKQAK--LRISVDFDELKEDPALEIHLLOHLKVAVACLDPRAMRPTWQVYMA 1154

DB 1018 VNNVRSYIRKDALSSGLVDAKRLLEBRIYSHMLTYIKIALCTSVSPVAPBMRQVILM 1077

QY 1155 FKEIQAGSG 1163

DB 1078 LIESERSEG 1086



Herbicide; plant; agriculture; herbicide.

Arabidopsis thaliana.

MO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(PARB) JAYER AG.

Tietjen K, Weidner M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms -

Claim 5; SEQ ID NO 249; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins (ABR9790-ABR94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.

Sequence 1133 AA;

Query Match 22.1%; Score 1367; DB 23; Length 1133;

Best Local Similarity 32.9%; Pred. No. 9.8e-95;

Matches 398; Conservative 179; Mismatches 455; Indels 176; Gaps 36;

8 PLSTVTLFFSFSSLPQASPSQSLYREIHQISFKVDLPDKNLLPDWSSNKNPCTFD 65

7 FLAIVILCSFSLIV-----RSINTEGRVLEFRKFLINDSNGYLAWNQDLSNPKWT 59

66 GUTCDDK-VTSIDLSKPLNVGSFSAVSSLSLTGLSLSFLNSHINGSV-SGFKCSAS 123

60 GIACHLRKRVTSVDLNGMLS---GTLSPICKLHGRKLVNSTVFISGPIPODSLGRS 116

124 LTLSDLSRNSLSG-----PYTTLTSL-----GSCGSLKPLNVSNLTD 161

117 LEVLDTCTRFHGVIPQLTMTITLTKLYLCENYLFSGIPROIGNLSLQELVITSSNL- 175

162 PFGKVSGL-KINSLEVLDSANISGANVGVLSGCGELKHLAISGNKISGV--DV 218

176 -TGVIPPSMAKRLQIIIRAGNGPSG--VI98EIS-GGESLKVIGLAKMLLEGLPQL 231

219 SRCVLEFLDVSSNNFSTGI-PFLGDCSALQHLDISGNKLSGDFSRASITCTELKILNIS 277

232 EKLQNTLDILWQNRSLSGEIPFVGNISRLLEVLAHENFTGSIPIREICKTKMRKLYLY 291

278 SNOFVGPIPL--PLKSLQYLSLAENKFTGERP-DPLSGACDTLGLDLSGNHFGAVP 334

292 TNOJTGEPREIGNLDIAEIDSENOJTGPIKEP--GHILMLKULHLFENILQPIPR 349

335 FFGSCSLLELSLSSNNFSGELPMDTLTKMRGKVLDSFNERSGELPSLTNLSASILT 394

350 ELGBLTLEKLDLSINRLNGTIPOE-LQFLPYVDLQFNOJLEKIP-PLIGFYENFSV 407

395 LDISNNFSGPIPLPNI.CQNKNTLQELYLQNGFTGKIPTTSLNCELYSIHLSFVYLSG 454

408 LDMSSANLSGPPIPAHFCR--FQTLILSLSGNLSGNIPDLTKCSLTETMLGDNQLTG 465

455 TIPSLSLSKLDKIDKLMLEGBIPOELMVYKLTETLIDENDLTGHPISGLSNCTNL 514

466 SLPIELFNQNLTALEHQLWLSGNISADIGKIKNLERLANNNTGHEIPPEIGULTKI 525

515 NWISLNNRLTGEIPKWIIGRLLENLAILKLSNNSPSGNIPELDCBSLWLIDLNTLFPNG 574

526 VGRNISSNQLTGHIPKELGSCVTIQLRDLISGNKFSGVIAQELQVLEYLETIRLSDRLTG 585

575 TIPAMFKSGKLANFIAKRYIYIKNDGKKECHQAGNLEFQGRSRBQLRLSTRNP 634

586 EIP-----HSFGDL-----TR-- 596

635 CNISRYVGGHTSPFPDNGSMFLDMSYMLSGYIPKEIGSMPLYFI-LNLGNDISGS 633

597 -----LMELQGLGNLSENIVEIGLTKLSIQISINISHNISGT 635

694 IPDEVGDLRGNTLIDLSNNKIDGRIPOAMGALTMULTIDLSNNNLSPGPIEMOGFEFFP 753

636 IPDSLGNLQMLEIYILNNDKLSGEIPASIGNLMSLI.CNLSNNNLVGTVPDPAVFQRMDS 695

754 AKFLANPGLCGYPLPRCDP--SNADGYAHHQSHGRBPASIA-----GSVAMGLRSPV 805

636 SNFAGNHGLCNSQSHCCPLVPHSDSKLMLNLSQOKILITICIVIGSV---FLITFL 752

806 CIFGLILVREKRRKRRKKEALEMTVAEGHNSGDRJANTNTMKTGVKALSTINLAPE 865

753 GLCWIT-----KRRERPAFALE-----DQKPDV-----MDSY 781

866 KPLRLTFADLIQATNGHNDLSIGSGFQDYVYKILKDSAAVAILKLIHVSQGG---DR 922

782 PFKGFYQGLVDAIRNPSSEVDVLRGACGVYVAEWSGGGVIAVKL-NRGRGASSDN 840

923 EMAMETIGIKHKNVLPVLYGCVGDERILVNEVMKYSLEBDVLPKXGKYLKLT 982

841 SFRASITLIGIKRHNVIKLYGFCYHONSLLIYVMSKSGLEQLGEGKNCI-LDWNA 899

983 RRKIAISARGAFLAHNCSPIHIDMKSSNVLDENLEAPVSPFGMARLMSADHTLS 1042

900 RRRIALGAEBGCIYLHDCRPQIVHRDIKSNLILDERFOAHVGFGLAKLID-LSYSKS 958

1043 VSTLAGTPGVPEPYQSRCTGKDVYSYGVILLELTGKRPDTPSPFGDNILVGVYKQ 1102

959 MSAVASGYIAPEYAVYTWKVKTEKDIYSFGVALLLELTGKRPVQPLDGG-CDLVNWR 1017

1103 HAKLRIS--DVEPPELMKDPALTEIELLOHLKVAAACLDRAARPPTVQVNAFKEIOA 1160

1018 STPNMIPTEMPARLDPTDKXTVHMSLVLKIALFCTNSPASRPMTREVVAVAMTEARG 1077

1161 GSGIDSGS 1168

1078 SSSLSSS 1085

RESULT 15

ABB91986

ID ABB91986 standard; Protein; 1124 AA.

XX ABB91986;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 1197.

XX Herbicide; plant; agriculture; herbicide.

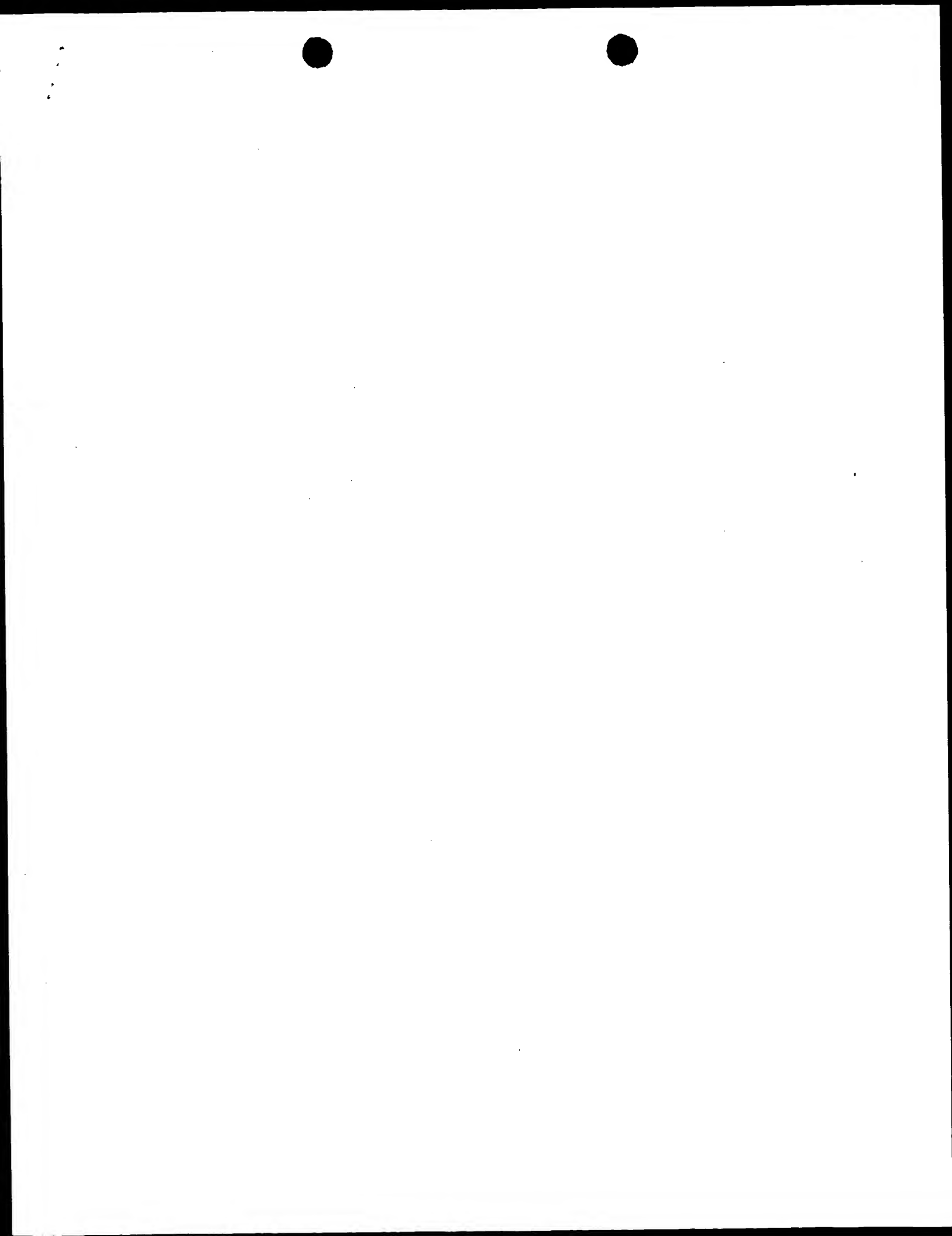
XX Arabidopsis thaliana.

XX MO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:02:09 ; Search time 2.85851 Seconds

(without alignments)  
4469.091 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649

Perfect score: 332

Sequence: 1 AANFIAKRVYIKNDGK.....STRPCNTSRVYGHSTPT 62

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	332	100.0	1196 10 022476	022476 arabidopsis
2	140.5	42.3	1121 10 0942P3	0942P3 oryza sativ
3	128	38.6	1166 10 092WC8	092WC8 arabidopsis
4	124	37.3	1166 10 09ARF3	09ARF3 capsella ru
5	116	34.9	1164 10 09LTF3	09LTF3 arabidopsis
6	115	34.6	1192 10 09ARC8	09ARC8 lycopersico
7	99.5	30.0	1110 10 094LN2	094LN2 oryza sativ
8	94.5	28.5	1143 10 09ZSS9	09ZSS9 arabidopsis
9	65	19.6	322 2 P941E8	0941E8 actinobacill
10	63.5	19.1	322 16 P941E8	0941E8 bruceella me
11	63.5	19.1	1436 5 0961S8	0961S8 plasmodium
12	63	19.0	916 17 09V1G8	09V1G8 pyrococcus
13	62.5	18.8	295 10 09MBG8	09MBG8 arabidopsis
14	62	18.7	871 17 059245	059245 pyrococcus
15	61	18.4	1592 5 09GUT7	09GUT7 lucilia cup
16	60.5	18.2	341 2 085214	085214 pseudomonas

17	60.5	18.2	466 2 0916M4	0916M4 pseudomonas
18	60.5	18.2	725 5 09VEG6	09VEG6 drosophila
19	60.5	18.2	1341 5 09VM09	09VM09 drosophila
20	60	18.1	1227 16 08ZOR1	08ZOR1 anabaena sp
21	60	18.1	3078 5 026031	026031 plasmodium
22	59.5	17.9	1195 5 09N343	09N343 caenorhabdi
23	59	17.8	333 16 09CHJ5	09CHJ5 lactococcus
24	59	17.8	408 16 09Z4S5	09Z4S5 salmoneilla
25	58.5	17.6	385 10 09FKX4	09FKX4 arabidopsis
26	58.5	17.6	776 17 027582	027582 methanobact
27	58.5	17.6	914 12 085427	085427 rat cytochrome
28	58	17.5	275 17 08TVD3	08TVD3 methanopyru
29	58	17.5	385 11 09R2D0	09R2D0 mus musculu
30	58	17.5	397 11 035325	035325 mus musculu
31	58	17.5	397 11 035118	035118 mus musculu
32	57.5	17.3	397 11 035118	035118 mus musculu
33	57.5	17.3	110 5 095X53	095X53 caenorhabdi
34	57.5	17.3	485 10 024665	024665 porphyra pu
35	57.5	17.3	523 10 09S7T6	09S7T6 arabidopsis
36	57.5	17.3	539 4 07S322	07S322 homo sapien
37	57.5	17.3	635 4 08TBA7	08TBA7 homo sapien
38	57.5	17.3	636 16 092N98	092N98 rhizobium m
39	57.5	17.3	733 11 091XW0	091XW0 ratuus norv
40	57.5	17.3	842 10 091UJ2	091UJ2 arabidopsis
41	57	17.2	107 16 092GN6	092GN6 rickettsia
42	57	17.2	126 6 09GMD0	09GMD0 macaca fasc
43	57	17.2	300 16 08UAX1	08UAX1 agrobacteri
44	57	17.2	386 4 0915R3	0915R3 homo sapien
45	57	17.2	1488 16 P96204	P96204 mycobacteri

## ALIGNMENTS

## RESULT 1

ID	022476	PRELIMINARY;	PRT; 1196 AA.
AC	022476;		
DT	01-JUN-1998 (TREMBLrel. 05, Created)		
DT	01-JUN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Brassinosteroid insensitive 1 (Brassinosteroid insensitive 1 gene) (BR11).		
GN	BR11 OR F23K16.30 OR AT4G39400.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OK	NCBI_TaxID=3702;		
RP	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COL-O;		
RX	MEDLINE=97442355; PubMed=9298904;		
RT	Li U., Chory J.;		
RT	"A putative leucine-rich repeat receptor kinase involved in		
RT	brassinosteroid signal transduction.";		
RL	Cell 90:929-938(1997).		
RL	(2)		
RP	SEQUENCE FROM N.A.		
RA	Beyan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,		
RA	Mayer K.F.X., Lemcke K., Schueller C.;		
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
RP	(3)		
RP	SEQUENCE FROM N.A.		
RA	EU Arabidopsis sequencing project;		
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.		
RP	(4)		
RP	SEQUENCE FROM N.A.		
RA	Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;		
RA	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RP	(5)		
RP	SEQUENCE FROM N.A.		
RA	EU Arabidopsis sequencing project;		



Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AF017056; AAC49810.1; -  
 DR EMBL; AL078620; CAB4675.1; -  
 DR EMBL; AL161595; CAB80603.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR Pfam; PF00560; LRR; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 15.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SEQUENCE 1196 AA; 130542 MW; C7FBA1C21294E600 CRC64;  
 Query Match 100.0%; Score 332; DB 10; Length 1196;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-33;  
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 AANFIAGKRYVYIKNDGKKECHGAGNLEFQGISQNLRLSTRNCPNITSRYGSHTS 60  
 Db 588 AANFIAGKRYVYIKNDGKKECHGAGNLEFQGISQNLRLSTRNCPNITSRYGSHTS 647  
 Oy 61 PT 62  
 Db 648 PT 649  
 RESULT 2  
 Q942F3 PRELIMINARY; PRT; 1121 AA.  
 ID Q942F3  
 AC 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Putative brassinosteroid-insensitive protein BRI1.  
 GN P0480C01.13.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NC NCB1\_TaxID=4530;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.,  
 RA "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0480C01."  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF003453; BAB68053.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00560; LRR; 19.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; UNKNOWN\_1.  
 KW ATP-binding; Transferase.  
 SO SEQUENCE 1121 AA; 120180 MW; F71A49B45E0E2D09 CRC64;  
 Query Match 42.3%; Score 140.5; DB 10; Length 1121;  
 Best Local Similarity 46.6%; Pred. No. 4.5e-09;  
 Matches 27; Conservative 11; Mismatches 19; Indels 1; Gaps 1;

Oy 5 IAGKRYVYIKNDGKKECHGAGNLEFQGISQNLRLSTRNCPNITSRYGSHTS 62  
 Db 520 IAGKRYVYIKNDGKKECHGAGNLEFQGISQNLRLSTRNCPNITSRYGSHTS 576  
 RESULT 3  
 Q92WC8 PRELIMINARY; PRT; 1166 AA.  
 ID Q92WC8  
 AC Q92WC8  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE F20N2.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NC NCB1\_TaxID=3702;  
 RP SEQUENCE FROM N.A.  
 RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,  
 RA Khan S., Alatafi H., Bel O., Chin C., Chio J., Choi E., Conn L.,  
 RA Shim P., Alatafi H., Bel O., Chin C., Chio J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vayberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RA "Genomic sequence for Arabidopsis thaliana BAC F20N2 from chromosome  
 I.1."  
 RT Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RA Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RA Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Shin P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,  
 RA Walker M.W., Alatafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,  
 RA Gonzalez A.A., Hansen N.N.F., Hutzar L.L., Kremetskaia I.I.,  
 RA Lenz C.C., Li J.J., Liu S.S., Luros S.S., Rowley D.D., Schwartz J.J.,  
 RA Toriumi M.M., Vayotskaya V.V., Yu G.G., Davis R.R.W.,  
 RA Toriumi M.M., Vayotskaya V.V., Yu G.G., Davis R.R.W.,  
 RA Federspiel N.N.A., Theologis A.A., Ecker J.J.R.;  
 RA Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RA Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Alatafi H., Bel B., Chin C., Chio J., Choi E.,  
 RA Khan L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vayberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RA Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AC002328; AAF79510.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR InterPro; IPR001245; TYR\_pkinase.  
 DR Pfam; PF00560; LRR; 21.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR PRINTS; PR00109; TYRKINASE.

DR ProDom: PD0000001; Euk\_pkinase; 1.  
 DR SMART: SM00370; LRR; 16.  
 DR SMART: SM00221; STYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KM ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 1166 AA; 127423 MW; 8C4D9231A466A7 CRC64;

Query Match 38.6%; Score 128; DB 10; Length 1166;  
 Best Local Similarity 43.1%; Pred. No. 1.9e-07;  
 Matches 25; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

OY 5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNCTITSRYVGHTSPT 62  
 Db 579 VSGKQFAFVFNNEG-GTDCRGAGGLVEFGIRARLERLHPMVHSCPAT-RISYGMTWT 634

RESULT 4

OC 09ARF3 PRELIMINARY; PRT; 1166 AA.  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical 127.3 kDa protein.  
 OS Capsella rubella.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Capella.  
 OC NCBI\_TaxID=81985;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21178822; PubMed=11283350;  
 RA Rosenberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,  
 RA Schumacher K., Schmitz G., Schmidt R.;  
 RT "Comparative sequence analysis reveals extensive microcolinearity in  
 RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
 RT genomes.";  
 RL Plant Cell 13:979-988(2001).  
 CC -1. SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AJ030349; CAC36390.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00560; LRR; 18.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; LEURICHRPT.  
 DR PRODOM; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KM ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;  
 Transferase.  
 SQ SEQUENCE 1166 AA; 127261 MW; DAD22B2B4C5647A0 CRC64;

Query Match 37.3%; Score 124; DB 10; Length 1166;  
 Best Local Similarity 41.4%; Pred. No. 6e-07;  
 Matches 24; Conservative 15; Mismatches 17; Indels 2; Gaps 2;

OY 5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNCTITSRYVGHTSPT 62  
 Db 579 VSGKQFAFVFNNEG-GTDCRGAGGLVEFGIRARLERLHPMVHSCPAT-RISYGMTWT 634

RESULT 5

OC 09LJF3 PRELIMINARY; PRT; 1164 AA.  
 AC 09LJF3;

DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Receptor protein kinase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.,  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA MEDLINE=20363099; PubMed=10907853;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,  
 RT TAC and BAC clones.";  
 RL DNA Res. 7:217-221(2000).  
 CC -1. SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AP000603; BAB01743.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00560; LRR; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; LEURICHRPT.  
 DR PRODOM; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 14.  
 DR SMART; SM00221; STYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KM ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 1164 AA; 126660 MW; 79380581D400EEEC CRC64;

Query Match 34.9%; Score 116; DB 10; Length 1164;  
 Best Local Similarity 40.0%; Pred. No. 6.3e-06;  
 Matches 22; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

OY 5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNCTITSRYVGHT 59  
 Db 579 VSGKQFAFVFNNEG-GTDCRGAGGLVEFGIRARLERLHPMVHSCPAT-RISYGMT 631

RESULT 6

OC 09ARC8 PRELIMINARY; PRT; 1192 AA.  
 ID 09ARC8;  
 AC 09ARC8;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical 129.9 kDa protein.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Asteridae; eussterids I; Solanales; Solanaceae; Solanum.  
 OC NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21178822; PubMed=11283350;  
 RA Rosenberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,  
 RA Schumacher K., Schmitz G., Schmidt R.;  
 RT "Comparative sequence analysis reveals extensive microcolinearity in  
 RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella

```

RT genomes."
CC Plant Cell 13:979-988(2001).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AJ303345; CAC36401.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00560; LRR; 21.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
DR ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
KW TRANSFERASE.
SEQUENCE 1192 AA; 129941 MW; BDEICDAP8930886 CRC64;

Query Match 34.6%; Score 115; DB 10; Length 1192;
Best Local Similarity 43.9%; Pred. No. 8.7e-06;
Matches 25; Conservative 12; Mismatches 18; Indels 2; Gaps 2;

Qy 6 AGKRVYIKNDGKKECHGAGNLEFQGISQNLRLSTRNPNITSRYVGGTSTPT 62
Db 609 SGKQPAFPANRG-GTECHGAGVGFEBIRERLALLPMVHFCST-RIYSGRTMYT 663

RESULT 7
Q94LN2 PRELIMINARY; PRT; 1110 AA.
ID Q94LN2
AC Q94LN2
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative receptor protein kinase.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriocarpaceae; Oryzaceae; Oryza.
OC NCBI_TaxID=4530;
OK [1]
RA WING R.A., Fritsch D., Presting G., Wood T., Yu Y., Soderlund C.,
RA Kim H., Rambo T., Henry D., Simmons J.;
RA Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC078891; AK5254.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00560; LRR; 19.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; UNKNOWN_1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
DR ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
KW TRANSFERASE.
SEQUENCE 1110 AA; 118109 MW; B6723380BC0A8E9A CRC64;

Query Match 30.0%; Score 99.5; DB 10; Length 1110;
Best Local Similarity 37.5%; Pred. No. 0.00076;
Matches 21; Conservative 11; Mismatches 21; Indels 3; Gaps 2;

Qy 2 ANFAGKRVYIKNDGKKECHGAGNLEFQGISQNLRLSTRNPNITSRYVGG 57
Db 530 SGILSGNTLAFVNRVG--NSCKVGGLLEFAGIRPERLLQVPTKSCDFT-RLVSG 582

```

```

RESULT 8
Q9ZPS9 PRELIMINARY; PRT; 1143 AA.
ID Q9ZPS9
AC Q9ZPS9
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative receptor protein kinase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
OK [1]
RA WING R.A., Fritsch D., Presting G., Wood T., Yu Y., Soderlund C.,
RA Kim H., Rambo T., Henry D., Simmons J.;
RA Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC065332; RAD20088.1; -.
DR EMBL; AY074313; AAL67010.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR Pfam; PF00560; LRR; 21.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00370; LRR; 18.
DR SMART; SM00221; STYKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
DR ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
KW TRANSFERASE.
SEQUENCE 1143 AA; 125675 MW; 7D1C88493F27A94E CRC64;

Query Match 28.5%; Score 94.5; DB 10; Length 1143;
Best Local Similarity 33.9%; Pred. No. 0.0034;
Matches 19; Conservative 13; Mismatches 21; Indels 3; Gaps 2;

Qy 2 ANFAGKRVYIKNDGKKECHGAGNLEFQGISQNLRLSTRNPNITSRYVGG 57

```

DB 549 SGLSGNTWAFVRNVG--NSCKGVGLVPSGIRPERLLQIPSLKSCDFT-RMTSG 601

## RESULT 9

ID P94168 PRELIMINARY; PRT; 322 AA.

AC P94168;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Mutarotase.

GN GALT.

OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;

OC Actinobacillus.

OC NCBI\_TaxID=715;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CM5;

R "galm gene of Actinobacillus pleuropneumoniae."

RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RA EMBL, U63731; AAB37129.1; -.

DR InterPro; IPR001823; Aldi epimerase.

DR Pfam; PF01263; Aldose epim; 1.

SO SEQUENCE 322 AA; 35349 MW; 0044BBD7F4777C5 CRC64;

Query Match 19.6%; Score 65; DB 2; Length 322;

Best Local Similarity 41.3%; Pred. No. 4.8;

Matches 19; Conservative 5; Mismatches 14; Indels 8; Gaps 2;

DB 60 ANRANAEYQNGKTYLVKNDG-KNTLHGANGADKQIWMQOLD 104

## RESULT 10

ID Q8YXH8 PRELIMINARY; PRT; 322 AA.

AC Q8YXH8;

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE L-asparaginase II (EC 3.5.1.1).

GN EMEI0105.

OS Brucella melitensis.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Brucellaceae; Brucella.

OC NCBI\_TaxID=29459;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;

RA MEDLINE=20020109; PubMed=11756688;

RA Delvecchio V.G., Kapriel V., Redkar R.J., Patra G., Mujar C., Los T.,

RA Ivanova N., Anderson I., Bhattacharya A., Lykide A., Reznik G.,

RA Jablonik L., Larsen N., D'Souza M., Bernal A., Mazur M., Goleman E.,

RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Ieeseon J.-J.,

RA Haselkorn R., Kyriades N., Overbeek R., Ieeseon J.-J.,

RT "The genome sequence of the facultative intracellular pathogen

RT Brucella melitensis."

RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).

RC EMBL, AE009453; AAL51287.1; -.

DR InterPro; IPR00267; Asp/Glutamase.

DR Pfam; PF00710; Asparaginase; 1.

DR PRINTS; PR00139; ASGNINASE.

DR Prodom; PD003221; Asp/Glutamase; 1.

DR Hydroxylase; Complete proteome.

SO SEQUENCE 322 AA; 34208 MW; 15A4638DB21C31B CRC64;

Query Match 19.1%; Score 63.5; DB 16; Length 322;

Best Local Similarity 33.3%; Pred. No. 7.5;

Matches 15; Conservative 10; Mismatches 17; Indels 3; Gaps 1;

OY 16 DMKKECHGAGNLEFGIRSEQLNRSTRNPNITRYGHTS 60

DB 241 DGIYVAGRGAGHV---SGDEADITERYASRIPVVASRTYGGRTA 282

## RESULT 11

ID O96158 PRELIMINARY; PRT; 1436 AA.

AC O96158;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Hypothetical 164.8 kDa protein.

GN PF0285C.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OC NCBI\_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=99021743; PubMed=9804551;

RA Gardner M.J., Tetelijn H., Carucci D.J., Cummings L.M., Aravind L.,

RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,

RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,

RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,

RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;

RT "Chromosome 2 sequence of the human malaria parasite Plasmodium

RT falciparum."

RL Science 282:1126-1132(1998).

DR EMBL; AE001385; AAC71846.1; -.

KW Hypothetical protein.

SO SEQUENCE 1436 AA; 164849 MW; 191BD374294E37EA CRC64;

Query Match 19.1%; Score 63.5; DB 5; Length 1436;

Best Local Similarity 33.3%; Pred. No. 39;

Matches 22; Conservative 10; Mismatches 19; Indels 15; Gaps 4;

OY 8 KRYVYIKNDG--MKKECHGAGNLEFGIRSEQ-----LNRSTRNPN-----ITSRV 54

DB 183 KAVSFKNDGIDINSKKNH--DNLMFLKIRSKSNMNLIVRKITHVNVNIVSGMTNKV 240

OY 55 YGHTS 60

DB 241 IGGWMS 246

## RESULT 12

ID O9V168 PRELIMINARY; PRT; 916 AA.

AC O9V168;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein PAB0383.

GN PAB0383.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OC NCBI\_TaxID=29292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORSAY;

RA Heilig R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome

RT structure and evolution."

RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

RA EMBL; AJ248284; CAB49482.1; -.

DR InterPro; IPR003586; Hedgehog\_hintc.

DR InterPro; IPR003587; Hedgehog\_hintc.

DR InterPro; IPR004042; Intein.

DR InterPro; IPR001233; UPF0027.

DR Pfam; PF01139; UPF0027; 2.

DR PRINTS; PR00379; INTEIN.

Search completed: March 10, 2005, 18:18:37  
Job time : 4.85851 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:16:35 ; Search time 0.985692 Seconds

(without alignments)  
2608.861 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649

Sequence: 1 AANFLAGKRYVYIKNDGKMK.....STRNFCNITSRVGHTSPT 62

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	18.4	313	1 PLDB_HAEMIN	P44800 haemophilus
2	58.5	17.6	316	1 YGUA_CAEEL	P30634 caenorhabdi
3	58	17.5	342	1 PG11_PHAUV	P35314 phaeocystis
4	58	17.5	2210	1 RRPL_EBOSM	O66892 ebola virus
5	57.5	17.3	719	1 HS9A_HORSE	O99X77 equus caball
6	57.5	17.3	728	1 HS9A_CHICK	P11501 gallus gall
7	57.5	17.3	731	1 HS9A_HUMAN	P07900 homo sapien
8	57.5	17.3	732	1 HS9A_CRIGR	P46633 cricetus
9	57.5	17.3	732	1 HS9A_MOUSE	P07901 mus musculu
10	57.5	17.3	732	1 HS9A_MOUSE	O02705 sus scrofa
11	57	17.2	342	1 PG13_PHAUV	P58823 phaeocystis
12	56	16.9	257	1 DLH1_SUISO	P95882 sulfolobus
13	56	16.9	337	1 G55A_CHICK	O98892 gallus gall
14	56	16.9	3462	1 RELN_RAT	P58751 rattus norv
15	55.5	16.7	118	1 MTPN_CHICK	O91955 gallus gall
16	55.5	16.7	658	1 TKT_STREN	P23976 streptococ
17	55	16.6	135	1 ES30_TRICO	O97391 crithosion
18	55	16.6	467	1 EMS5_FUGRU	P49697 fuquibacter
19	55	16.6	474	1 Y068_MTCGE	P43714 mycoplasma
20	55	16.6	1064	1 NFC2_MOUSE	O60591 mus musculu
21	54.5	16.4	383	1 DNUA_LACSK	O87782 lactobacill
22	54.5	16.4	384	1 PQOE_ACICA	P07782 acinetobact
23	54.5	16.4	342	1 PG12_MOUSE	O06601 comamonas
24	54	16.3	342	1 EXTL_CLOAB	P58822 phaeocystis
25	54	16.3	399	1 EXTL_CLOAB	O97900 clostridium
26	53.5	16.1	278	1 T2P2_NEIGO	P24617 neisseria
27	53.5	16.1	1073	1 MTR4_YEAST	P47047 saccharomy
28	53	16.0	345	1 OPCM_RAT	P32736 rattus norv
29	53	16.0	432	1 AP2_ARATH	P47927 arabidopsis
30	53	16.0	706	1 YX14_CAEEL	O11125 caenorhabdi
31	52.5	15.8	112	1 RL34_SCHPO	O42946 schizosacch
32	52.5	15.8	433	1 KARI_YEAST	P11927 saccharomy
33	52	15.7	198	1 YMO3_YEAST	O03673 saccharomy

34	52	15.7	229	1 C79B_HUMAN	P40259 homo sapien
35	52	15.7	401	1 KYB1_CHICK	O99W1 gallus gall
36	52	15.7	710	1 PPEC_BOVIN	O97A2 bos taurus
37	51.5	15.5	216	1 PUR7_AQUAE	O67881 aquifex aeo
38	51.5	15.5	227	1 YWMC_BACSU	P70960 bacillus au
39	51.5	15.5	294	1 PANE_ARCFU	O28578 archaeoglob
40	51.5	15.5	324	1 GLXA_RHIME	O87389 rhizobium m
41	51.5	15.5	1284	1 ART_COMPLEX	P16602 complex viru
42	51	15.4	203	1 YME6_YEAST	P38428 saccharomy
43	51	15.4	324	1 DLH1_CANAL	P50265 candida alb
44	51	15.4	345	1 OPCM_HUMAN	O14982 homo sapien
45	51	15.4	392	1 SB11_HUMAN	O96p15 homo sapien

## ALIGNMENTS

RESULT 1	PLDB_HAEMIN	STANDARD	PRT	313 AA.
AC	P44800:			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DE	15-UN-2002 (Rel. 41, Last annotation update)			
DE	Probable lysophospholipase L2 (EC 3.1.1.5) (lecithinase B).			
OS	PLDB OR H10645.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus.			
OK	NCBI_TaxID=727;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=RD / KM20 / ATCC 51907;			
RX	MEDLINE=95350630, PubMed=7542801;			
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness B.F.,			
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,			
RA	Weidman J.F., Phillips C.A., Spriggs C.A., Kellley J.M.,			
RA	Uterback T.R., Hanna M.C., Nguyen D.T., Sadek D.M., Brandon R.C.,			
RA	Pine L.D., Fritchman J.L., Fuhrman J.L., Geoghegan N.S.M.,			
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,			
RA	Venter J.C.;			
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae			
RT	Rd.";			
RL	Science 269:496-512 (1995).			
CC	-1- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =			
CC	glycerophosphocholine + a fatty acid anion.			
CC	-1- SUBCELLULAR LOCATION: Inner membrane (By similarity).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
DR	EMBL; U37747; AAC2305.1; -			
DR	TIGR; H10645; -			
DR	InterPro; IPR000073; Abhydrolase.			
DR	InterPro; IPR000379; Ser_estr_ site.			
DR	Pfam; PF00561; abhydrolase; 1.			
KW	Hydrolase; Lipid synthesis; Inner membrane; Complete proteome.			
SQ	SEQUENCE 313 AA; 36656 MW; 891E784FF77C28 CRC64;			
Query Match	18.4%; Score 61; DB 1; Length 313;			
Best Local Similarity	36.6%; Pred. No. 1.6;			
Matches 15; Conservative	9; Mismatches 15; Indels 2; Gaps 1;			
QY	7 GKRYYVYIKNDGKKEGAGNLLFPGIRSEOLNLSRNP 47			
Db	181 GERYVFGK-GAYCOAHLEVNELTFCKTRMKMNRINKRNP 219			

RESULT 2  
YOUNG CAEEL STANDARD; PRT; 316 AA.  
AC P30634;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
GN Hypothetical 36.5 kDa protein ZK637.11 in chromosome III.  
OS *Caenorhabditis elegans*.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; *Caenorhabditis*.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Br16161; PubMed=1538779;  
RX MEDLINE=92168156; PubMed=1538779;  
RA Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,  
Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,  
Craxson M., Durbin R., Barks M., Metzstein M., Hawkins T.,  
Ainscough R., Waterston R.;  
RT "The C. elegans genome sequencing project: a beginning."  
RL Nature 356:37-41 (1992).  
CC -1- SIMILARITY: BELONGS TO THE MPI PHOSPHATASE FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 RHODANES DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, Z11115; CAA77456.1; -.  
DR PIR, S15799; S15799.  
DR HSP, P30305; 10B0.  
DR WormPep; ZK637.11; CE00429.  
DR InterPro; IPR001763; Rhodanese-like.  
DR Pfam; PF00581; Rhodanese, 1.  
DR SMART; SM00450; RHOD; 1.  
DR Hypothetical protein.  
KW SEQUENCE 316 AA; 36469 MW; 6C6EA2921623DB59 CRC64;  
Query March 17.6%; Score 58.5; DB 1; Length 316;  
Local Similarity 27.3%; Pred. No. 3.4;  
Ches 15; Conservative 13; Mismatches 22; Indels 5; Gaps 2;  
CY 13 INDDMKKRC---HGAGNLEFGIR-SEQLNRLSTNPNCTITRYVGHSTPT 62  
DB 12 VANDGLRLKSCSACGSSKLPFRQNRHSSAISHTSSPPTRKRSIDGGYTSST 66  
RESULT 3  
PGII PHAVU STANDARD; PRT; 342 AA.  
AC P35334;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
GN Polygalacturonase-inhibitor 1 precursor (Polygalacturonase-inhibiting  
DE protein) (PGIP-1).  
OS *Phaseolus vulgaris* (Kidney bean) (French bean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
OX NCBI\_TaxID=3885;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=cv, Saxa; TISSUE=Hypocotyl;

RX MEDLINE=93272053; PubMed=1303801;  
RA Touchart P., Desiderio A., Salvi G., Ceryone F., Daroda L.,  
de Lorenzo G., Bergmann C., Darvill A.G., Albersheim P.;  
RT "Cloning and characterization of the gene encoding the  
RT endopolygalacturonase-inhibiting protein (PGIP) of *Phaseolus vulgaris*  
RT L.";  
RL Plant J. 2:367-373 (1992).  
RN [2]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF LYS-253.  
RC STRAIN=cv, Pinto; TISSUE=Hypocotyl;  
RX MEDLINE=9246261; PubMed=10228150;  
RA Leckie F., Mattei B., Capodice C., Hemmings A., Nuss L., Atacri B.,  
De Lorenzo G., Ceryone F.;  
RT "The specificity of polygalacturonase-inhibiting protein (PGIP): a  
RT single amino acid substitution in the solvent-exposed  
RT beta-strand/beta-turn region of the leucine-rich repeats (LRRs)  
RT confers a new recognition capability.";  
RL EMBO J. 18:2352-2363 (1999).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=97449843; PubMed=9304859;  
RA Desiderio A., Atacri B., Leckie F., Mattei B., Salvi G., Tigelaar H.,  
Van Roekel J.S., Baulcombe D.C., Melchers L.S., De Lorenzo G.,  
Ceryone F.;  
RT "Polygalacturonase-inhibiting proteins (PGIPs) with different  
RT specificities are expressed in *Phaseolus vulgaris*."  
RL Mol. Plant Microbe Interact. 10:852-860 (1997).  
CC -1- FUNCTION: Inhibitor of fungal polygalacturonase. It is an  
CC important factor for plant resistance to phytopathogenic fungi.  
CC Substrate preference is polygalacturonase (PG) from *A. niger* >> PG  
CC of *F. oxysporum*, *A. solani* or *B. cinerea*. Not active on PG from  
CC *F. moniliforme*.  
CC -1- SUBCELLULAR LOCATION: Cell-wall associated.  
CC -1- MISCELLANEOUS: Mutation of Lys-253 confers the ability to inhibit  
CC the *F. moniliforme* PG.  
CC -1- SIMILARITY: BELONGS TO THE POLYGALACTURONASE-INHIBITING PROTEIN  
CC FAMILY.  
CC -1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).  
CC -1- CAUTION: It is uncertain whether Met-1 or Met-10 is the initiator.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, X64769; CAA46016.1; -.  
DR EMBL, A23205; CAA01664.1; -.  
DR PIR, S23764; S23764.  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR003592; LRR\_out.  
DR Pfam; PF00560; LRR; 5.  
DR SMART; SM00370; LRR; 5.  
KW Signal; Glycoprotein; Cell wall; Repeat; Leucine-rich repeat.  
FT SIGNAL 1 29  
FT CHAIN 30 342  
FT REPEAT 78 104  
FT REPEAT 129 152  
FT REPEAT 153 177  
FT REPEAT 179 200  
FT REPEAT 201 225  
FT REPEAT 272 296  
FT REPEAT 298 321  
FT DISULFID 32 62  
FT DISULFID 63 72  
FT DISULFID 310 332  
FT DISULFID 334 341  
FT CARBOHYD 64 64  
FT CARBOHYD 141 141  
FT CARBOHYD 303 303  
FT MOTTGN 253 253  
POLYGALACTURONASE INHIBITOR 1.  
LRR 1.  
LRR 2.  
LRR 3.  
LRR 4.  
LRR 5.  
LRR 6.  
LRR 7.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
K->Q: BROADER SPECTRUM OF ACTION.



SEQUENCE 342 AA; 37101 MW; 950F94E0DD2A39598 CRC64;

Query Match 17.5%; Score 58; DB 1; Length 342;  
Best Local Similarity 34.5%; Pred. No. 4.3;  
Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;

3 NFINGKRVYIKNDGKKECHGAGNLEFQ---GISEQLNRSTNPNCTTSRYVG 56  
Db 236 NMEGDSVLFSGDKNTKTHLAKNSLAFLDGKVGTL-SKINLGDLENN-----NRIYG 287

## RESULT 4

RRPL\_EBOSM STANDARD; PRT; 2210 AA.

AC 066802;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein)  
CC (L protein).

OC Ebola virus (strain Sudan Maleo-79) (Bbo).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;  
OC Ebola-like viruses.

OX NCBI\_TaxID=128949;

[1]

SEQUENCE FROM N.A.

RA Sanchez A., Trappler S., Nichol S.T.,  
Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACTIVE POLYMERASE. IT  
MAY FUNCTION IN RNA SYNTHESIS, CAPING, AS WELL AS METHYLATION OF  
CAPS. AND POLY(A) SYNTHESIS. ALSO INVOLVED IN TRANSCRIPTIONAL RNA  
EDITING.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA) (N).

-1- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS  
OF RHABDOVIRUSES AND PARAMYXOVIRUSES.

CC The SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U23458; AAA7970.1; -

DR InterPro; IPR001016; Viral\_RNA\_pol\_L.

DR Pfam; PF00946; Paramyx\_RNA\_pol\_1.

DR Ktransferrase; RNA-directed RNA polymerase.

SQ SEQUENCE 2210 AA; 251255 MW; F50EBB624951AER4 CRC64;

Query Match 17.5%; Score 58; DB 1; Length 2210;  
Best Local Similarity 39.5%; Pred. No. 34;  
Matches 17; Conservative 8; Mismatches 14; Indels 4; Gaps 2;

21 ECHGAGNLEFQ--GIRSEQLNRSTNPNCTTSRYGHTSP 61  
Db 1832 BEEGGALLIIOKYGKPLNTLATER--STESVIGSYTP 1872

RESULT 5

HS9A\_HORSE STANDARD; PRT; 719 AA.

ID HS9A\_HORSE STANDARD; PRT; 719 AA.  
AC 09GKX7;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Heat shock protein HSP 90-alpha (HSP 86) (Fragment).  
GN HSPCA.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

OX NCBI\_TaxID=9796;

BN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=21155861; PubMed=11258446;

RT "Molecular cloning of horse Hsp90 cDNA and its comparative analysis

RT with other vertebrate Hsp90 sequences."

RL J. Vet. Med. Sci. 63:115-124(2001).

CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY

CC (BY SIMILARITY).

CC -1- SUBUNIT: Homodimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AB043677; BAB2077.1; -

DR HSSP; P07900; IBO.

DR InterPro; IPR003594; ATPbind\_ATPase.

DR InterPro; IPR001404; Hsp90.

DR Pfam; PF00183; HSP90; 1.

DR Pfam; PF02518; HATPase\_C; 1.

DR PRINTS; PR00775; HEATSHOCK90.

DR SMART; SM00387; HATPase\_C; 1.

DR PROSITE; PS00298; HSP90; 1.

KW Chaperone; ATP-binding; Heat shock; Phosphorylation.

FT NON\_RES 1 1

FT MOD\_RES 220 220

FT MOD\_RES 252 252

FT MOD\_RES 719 719

FT NON\_TER

SQ SEQUENCE 719 AA; 83098 MW; EDCB1EE439FAC44D CRC64;

Query Match 17.3%; Score 57.5; DB 1; Length 719;  
Best Local Similarity 37.0%; Pred. No. 11;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

14 KNDGKKECHGAGNLEFQIRSEQLNR--LSTRNPNCTTSRYG 56  
Db 259 KKGDKKK---KKIKKXYIDGELNKTPIWTRPDDITNEYG 300

RESULT 6

HS9A\_CHICK STANDARD; PRT; 728 AA.

ID HS9A\_CHICK STANDARD; PRT; 728 AA.  
AC P11501;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Heat shock protein HSP 90-alpha.  
GN HSPCA.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Stomach, and Gizzard;  
RX MEDLINE=89165846; PubMed=2923621;  
RA Binart N., Chambraud B., Dumas B., Rowlands D.A., Bigogne C.,  
Levin J.M., Gantier J., Baulieu E.E., Catelli M.G.,  
"The cDNA-derived amino acid sequence of chick heat shock protein Mr  
90,000 (HSP 90) reveals a 'DNA like' structure: potential site of  
interaction with steroid receptors."  
RL Biochem. Biophys. Res. Commun. 159:140-147(1989).  
RN [2]  
RP SEQUENCE OF 1-85 FROM N.A.

RX MEDLINE=89345085; PubMed=2762125;  
 RA Vourc'H C., Binart N., Chambraud B., David J.P., Jerome V.,  
 RA Baillet E.E., Catelli M.G.,  
 RT "Isolation and functional analysis of chicken 90-kDa heat shock  
 RT protein gene promoter."  
 RL Nucleic Acids Res. 17:5259-5272(1989).  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 D EMBL; X07265; CAA30251.1; -;  
 D EMBL; X15028; CAA3132.1; -;  
 D PIR; S10880; HCH90.  
 DR HSP; P07900; IBYQ.  
 DR InterPro; IPR003594; ATPbind\_ATPase.  
 DR InterPro; IPR01404; Hsp90.  
 DR Pfam; PF00183; HSP90; 1.  
 DR Pfam; PF02518; HATPase\_C; 1.  
 DR PRINTS; PR00775; HEATSHOCK90.  
 DR SMART; SM00387; HATPase\_C; 1.  
 DR PROSITE; PS00298; HSP90; 1.  
 DR Chapter; ATP-binding; Heat shock; Phosphorylation.  
 DR MOD\_RES; 230 230 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES; 259 259 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 728 AA; 84059 MW; 6EC5660F4D5006CE CRC64;  
 Query Match 17.3%; Score 57.5; DB 1; Length 728;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
 Oy 14 KNDGKKECHGAGNLEFGIRSEQLNR---LSTNPNCNITSRYVG 56  
 Db 265 KNDGKKECHGAGNLEFGIRSEQLNR---LSTNPNCNITSRYVG 306  
 RESULT 7  
 HS9A\_HUMAN STANDARD; PRT; 731 AA.  
 ID HS9A\_HUMAN  
 UN P07900; OSBV05;  
 UN 01-AUG-1988 (Rel. 08, Created)  
 UN 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Heat shock protein HSP 90-alpha (HSP 86).  
 GN HSP90A OR HSP91 OR HSP90A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Peripheral blood lymphocytes;  
 RX MEDLINE=89386066; PubMed=2780322;  
 RA Soeda E., Yokoyama K., Yamazaki M., Akaogi K., Miwa T., Imai T.;  
 RT "Nucleotide sequence of a full-length cDNA for 90 kDa heat-shock  
 RT protein from human peripheral blood lymphocytes."  
 RL Nucleic Acids Res. 17:7108-7108(1989).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=91242090; PubMed=1368637;  
 RA Yamazaki M., Tashiro H., Yokoyama K., Soeda E.;  
 RT "Molecular cloning of cDNA encoding a human heat-shock protein whose  
 RT expression is induced by adenovirus type 12 E1A in HeLa cells."  
 RL Agric. Biol. Chem. 54:3163-3170(1990).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=89343979; PubMed=2527334;  
 RA Hickey E., Brandon S.E., Smale G., Lloyd D., Weber L.A.;  
 RT "Sequence and regulation of a gene encoding a human 89-Kilodalton  
 RT heat shock protein."  
 RL Mol. Cell. Biol. 9:2615-2626(1989).  
 RN [14]  
 RP SEQUENCE OF 1-311 FROM N.A.  
 RX MEDLINE=88056312; PubMed=2445630;  
 RA Hoffmann T., Hovemann B.;  
 RT "Cloning and nucleotide sequence of the murine hsp94 cDNA and  
 RT chromosome assignment of related sequences."  
 RL Gene 56:29-40(1987).  
 RN [15]  
 RP SEQUENCE OF 1-311 FROM N.A.  
 RX MEDLINE=90076956; PubMed=2591742;  
 RA Walter T., Drabant B., Krebs H., Tomalak M., Heiss S.,  
 RA Benecke B.J.J.;  
 RT "Cloning and analysis of a human 86-kDa heat-shock-protein-encoding  
 RT gene."  
 RL Gene 83:105-115(1989).  
 RN [16]  
 RP SEQUENCE OF 184-731 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Straubeberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [17]  
 RP SEQUENCE OF 538-731 FROM N.A.  
 RC TISSUE=Heart;  
 RA Tanaka M., Tanaka T., Mitani Y., Yamamoto M., Wood J.N.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [18]  
 RP SEQUENCE OF 1-20 AND PHOSPHORYLATION.  
 RX MEDLINE=89123325; PubMed=2492519;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "Two human 90-kDa heat shock proteins are phosphorylated in vivo at  
 RT conserved serines that are phosphorylated in vitro by casein kinase  
 RT II."  
 RL J. Biol. Chem. 264:2431-2437(1989).  
 RN [19]  
 RP PHOSPHORYLATION BY DS-DNA KINASE.  
 RX MEDLINE=9008887; PubMed=2507541;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "The human double-stranded DNA-activated protein kinase phosphorylates  
 RT the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal  
 RT threonine residues."  
 RL J. Biol. Chem. 264:17275-17280(1989).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 10-222.  
 RX MEDLINE=97262065; PubMed=9108479;  
 RA Stebbins C.E., Russo A.A., Schneider C., Rosen N., Hartl F.U.,  
 RA Pavletich N.P.;  
 RT "Crystal structure of an Hsp90-geldanamycin complex: targeting of a  
 RT protein chaperone by an antitumor agent."  
 RL Cell 89:239-250(1997).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 10-222.  
 RX MEDLINE=99034582; PubMed=9817749;  
 RA Obermann W.M., Sonderegger H., Russo A.A., Pavletich N.P., Hartl F.U.;  
 RT "In vivo function of Hsp90 is dependent on ATP binding and ATP  
 RT hydrolysis."  
 RL J. Cell Biol. 143:901-910(1998).  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X15183; CA33259.1; -  
 DR EMBL; X07270; CA30255.1; -  
 DR EMBL; M27024; AAA63194.1; -  
 DR EMBL; M30626; AAA63023.1; -  
 DR EMBL; BC000987; AAH00987.1; -  
 DR EMBL; D87666; BAA13430.1; -  
 DR EMBL; D87666; BAA13431.1; -  
 DR PIR; A32319; HHHU86.  
 DR PIR; J00724; B31420.  
 DR PIR; J00724; J00724.  
 DR PDB; 1YER; 22-APR-98.  
 DR PDB; 1YER; 22-APR-98.  
 DR PDB; 1YER; 22-APR-98.  
 DR PDB; 1YER; 22-APR-98.  
 DR TRANSFAC; T00992; -  
 DR Gene; HGNC:5253; HSPCA.  
 DR EMBL; X140571; -  
 DR InterPro; IPR003594; ATPbind\_ATPase.  
 DR InterPro; IPR001404; Hsp90.  
 DR Pfam; PF00183; HSP90; 2.  
 DR Pfam; PF02518; HATPase\_C; 1.  
 DR PRINTS; PR00775; HEATSHOCK90.  
 DR SMART; SM00387; HATPase\_C; 1.  
 DR PROSITE; PS00298; HSP90; 1.  
 DR Chaperone; ATP-binding; Heat shock; Phosphorylation; 3D-structure.  
 KW INIT\_MET 0 0  
 FT MOD\_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD\_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD\_RES 230 230 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD\_RES 262 262 PHOSPHORYLATION.  
 FT MOD\_RES 262 262 PHOSPHORYLATION.  
 FT CONFLICT 62 62 T -> S (IN REF. 3, 4 AND 5).  
 SQ SEQUENCE 731 AA; 84542 MW; 1249ABCFCE06297C CRC64;

Query Match 17.3%; Score 57.5; DB 1; Length 731;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGIRSEQLNR---LSTRNPNITSRVYG 56  
 DB 268 KKGDKKK---KKKIKKEYIDQELNKTPIWTNPNPDITNEEYG 309

RESULT 8  
 CRIGR STANDARD; PRT; 732 AA.  
 HSPA\_MOUSE  
 P46633;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-1995 (Rel. 34, Last annotation update)  
 DE Heat shock protein HSP 90-alpha (HSP 86).  
 GN HSPCA OR HSP90A.  
 OS Citricellus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 OK NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chen M.S.M.C., Laszlo A.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L33676; AAA6992.1; -  
 DR HSP; P07900; 1BYO.  
 DR InterPro; IPR003594; ATPbind\_ATPase.  
 DR InterPro; IPR001404; Hsp90.  
 DR Pfam; PF00183; HSP90; 1.  
 DR Pfam; PF02518; HATPase\_C; 1.  
 DR PRINTS; PR00775; HEATSHOCK90.  
 DR SMART; SM00387; HATPase\_C; 1.  
 DR PROSITE; PS00298; HSP90; 1.  
 DR Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 KW INIT\_MET 0 0  
 FT MOD\_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD\_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD\_RES 230 230 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD\_RES 262 262 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 732 AA; 84717 MW; AAF118902FAB3402 CRC64;

Query Match 17.3%; Score 57.5; DB 1; Length 732;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGIRSEQLNR---LSTRNPNITSRVYG 56  
 DB 269 KKGDKKK---KKKIKKEYIDQELNKTPIWTNPNPDITNEEYG 310

RESULT 9  
 HSPA\_MOUSE  
 ID HSPA\_MOUSE STANDARD; PRT; 732 AA.  
 AC P07901;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Heat shock protein HSP 90-alpha (HSP 86) (Tumor specific  
 DE transplacentalion 86 kDa antigen) (TSTA).  
 GN HSPCA OR HSP86-1 OR HSP86.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OK NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c;  
 RX MEDLINE=89174568; PubMed=2925609;  
 RA Moore S.K., Kozak C., Robinson E.A., Ullrich S.J., Appella E.;  
 RT "Murine 86- and 84-kDa heat shock proteins, cDNA sequences,  
 RT chromosome assignments, and evolutionary origins."  
 RT J. Biol. Chem. 264:5343-5351(1989).  
 RN [2]  
 RP SEQUENCE OF 5-355 FROM N.A.  
 RX MEDLINE=89232740; PubMed=2469626;  
 RA Hoffmann T., Hovemann B.;  
 RT "Heat-shock proteins, Hsp84 and Hsp86, of mice and men: two related  
 RT genes encode formerly identified tumour-specific transplacentalion  
 RT antigens."  
 RT Gene 74:491-501(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92009901; PubMed=1916807;  
 RA Moore S.K., Appella E., Villar C.U., Kozak C.A.;  
 RT "Mapping of the mouse 86-kDa heat-shock protein expressed gene  
 RT (Hsp86-1) on chromosome 12 and related genes on chromosomes 3, 4, 9,  
 RT and 11."  
 RT Genomics 10:1019-1029(1991).  
 RN [4]  
 RP SEQUENCE OF 459-732 FROM N.A.

RX MEDLINE=90033873; PubMed=2806771;  
 RA LeGagneux V., Megey V., Quelard C., Barnier J.V., Bensaud O.,  
 RA Morange M.;  
 RT "High constitutive transcription of HSP6 gene in murine embryonal  
 carcinoma cells";  
 RT Differentiation 41:42-48(1989).  
 RN  
 RP SEQUENCE OF 1-30.  
 RX MEDLINE=86205848; PubMed=3458168;  
 RA Ullrich S.J., Robinson E.A., Law L.W., Willingham M., Appella E.;  
 RT "A mouse tumor-specific transplantation antigen is a heat shock-  
 related protein";  
 RT Proc. Natl. Acad. Sci. U.S.A. 83:3121-3125(1986).  
 RN  
 RP PHOSPHORYLATION BY DS-DNA KINASE.  
 RX MEDLINE=90008887; PubMed=2507541;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "The human double-stranded DNA-activated protein kinase phosphorylates  
 the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal  
 threonine residues";  
 RT J. Biol. Chem. 264:17275-17280(1989).  
 CC  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL: U04633; AAA53068.1; -; ALT\_SEQ.  
 DR EMBL: M36830; AAA37868.1; ALT\_SEQ.  
 DR EMBL: M57673; AAA37867.1; -;  
 DR EMBL: X16857; CAA34748.1; -;  
 DR PIR: B32848; HHMS86.  
 DR HSP: P07900; IRYO.  
 DR MGD: MGI:96250; Hsp86-1.  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR001404; Hsp90.  
 DR Pfam: PF00183; HSP90.1.  
 DR Pfam: PF02518; HATPase\_c.1.  
 DR PRINTS: PR00775; HEATSHOCK90.  
 DR SMART: SM00387; HATPase\_c.1.  
 DR PROSITE: PS00298; HSP90.1.  
 DR Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT INIT MET 0 0  
 FT MOD RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 230 230 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD RES 262 262 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD RES 242 245 T->A (IN REF. 2).  
 FT CONFLICT 242 245 MISSING (IN REF. 2).  
 FT CONFLICT 355 355 R->K (IN REF. 2).  
 SO SEQUENCE 732 AA; 84656 MW; E6345DEBADAQCF1B CRC64;  
 Query Match 17.3%; Score 57.5; DB 1; Length 732;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
 Oy 14 KNDGKKECHGAGNLEFQGISRSEQLR--LSTRNPNITSRVYG 56  
 Db 269 KKGDKKK---KKTKKXYIDQELNKTPTWTRNPDITNEYG 310  
 RESULT 10  
 ID HS9A\_PIG STANDARD; PRT; 732 AA.  
 AC 002705;

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Heat shock protein HSP 90-alpha (HSP 86).  
 GN HSP90 OR HSP90A.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxID=9823;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99132303; PubMed=9931505;  
 RA Huang H.W., Lee W.C., Lin J.H., Ulan S.C., Mao S.J., Yang P.C.,  
 RA Huang T.Y., Liu Y.C.;  
 RT "Molecular cloning and characterization of porcine cDNA encoding a  
 90-kDa heat shock protein and its expression following  
 hyperthermia";  
 RT Gene 226:307-315(1999).  
 CC  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: Preferentially expressed in pituitary gland,  
 brain, adrenal gland, and testis, in comparison to other tissues.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL: U94395; AAC8718.1; -;  
 DR HSP: P07900; IRYO.  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR001404; Hsp90.  
 DR Pfam: PF00183; HSP90.1.  
 DR Pfam: PF02518; HATPase\_c.1.  
 DR PRINTS: PR00775; HEATSHOCK90.  
 DR SMART: SM00387; HATPase\_c.1.  
 DR PROSITE: PS00298; HSP90.1.  
 DR Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT INIT MET 0 0  
 FT MOD RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 6 6 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD RES 230 230 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 262 262 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 732 AA; 84643 MW; 759850062F4E5490 CRC64;  
 Query Match 17.3%; Score 57.5; DB 1; Length 732;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
 Oy 14 KNDGKKECHGAGNLEFQGISRSEQLR--LSTRNPNITSRVYG 56  
 Db 269 KKGDKKK---KKTKKXYIDQELNKTPTWTRNPDITNEYG 310  
 RESULT 11  
 ID PG13\_PHAVU STANDARD; PRT; 342 AA.  
 AC P58823;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Polylacturonase inhibitor 3 precursor (Polylacturonase-inhibiting  
 protein) (PGIP-2) (PGIP-3).  
 GN PGIP3.

OS Phaseolus vulgaris (Kidney bean) (French bean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Ericales; Fabiales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
 OK NCBI\_TaxID=3885;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=cv. Pinto; TISSUE=HYPOCOTYL;  
 RX MEDLINE=93272053; PubMed=1303801;  
 RA Toubart P., Desiderio A., Salvi G., Cervone F., Daroda L.,  
 de Lorenzo G., Bergmann C., Davylli A.G., Alberheim P.,  
 RT "Cloning and characterization of the gene encoding the  
 RT endopolygalacturonase-inhibiting protein (PGIP) of Phaseolus vulgaris  
 RT L.",  
 RL Plant J. 2:367-373(1992).  
 CC -1- FUNCTION: Inhibitor of fungal polygalacturonase. It is an  
 CC important factor for plant resistance to phytopathogenic fungi.  
 CC -1- SUBCELLULAR LOCATION: Cell-wall associated.  
 CC -1- TISSUE SPECIFICITY: Found in suspension-cultured cells and to a  
 CC lesser extent in hypocotyls, leaves and flowers.  
 CC -1- SIMILARITY: BELONGS TO THE POLYGALACTURONASE-INHIBITING PROTEIN  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).  
 CC -1- CAUTION: It is uncertain whether Met-1 or Met-10 is the initiator.  
 DR InterPro: IPR001611; LRR.  
 DR Pfam: PF00560; LRR\_5.  
 DR SMART: SM00370; LRR; 5.  
 KW Signal; Glycoprotein; Cell wall; Repeat; Leucine-rich repeat.  
 FT CHAIN 1 29  
 FT REPEAT 30 342 POLYGALACTURONASE INHIBITOR 3.  
 FT REPEAT 78 104 LRR 1.  
 FT REPEAT 129 152 LRR 2.  
 FT REPEAT 153 177 LRR 3.  
 FT REPEAT 179 200 LRR 4.  
 FT REPEAT 201 225 LRR 5.  
 FT REPEAT 226 256 LRR 6.  
 FT REPEAT 257 298 LRR 7.  
 FT REPEAT 299 321 LRR 8.  
 FT DISULFID 32 62 BY SIMILARITY.  
 FT DISULFID 63 72 BY SIMILARITY.  
 FT DISULFID 310 332 BY SIMILARITY.  
 FT DISULFID 334 341 BY SIMILARITY.  
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 342 AA; 37251 MW; 980AEP1DA28F27D CRC64;  
 QY Query Match 17.2%; Score 57; DB 1; Length 342;  
 DB Local Similarity 32.8%; Pred. No. 5.8;  
 DB Local 19; Conservative 9; Mismatches 20; Indels 10; Gaps 3;  
 DB 236 MNLQGDASTVPSGDNTQKIKLAKNSLDFLEKVGSL-SKINGLDLRN-----NRIYV 287

RESULT 12  
 ID DLH SULSO STANDARD; PRT; 257 AA.  
 AC P95862;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative carboxymethylglutaminylase (Ec 3.1.1.45) (Dienelactone  
 DE hydrolase) [DLH].  
 GN SSO2087 OR C06015 OR G31\_045.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OK NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;

RX MEDLINE=97055432; PubMed=8899719;  
 RA Sersen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,  
 RA Liu O.Y., Penny S.L., Young F., Schenk W.E., Gaasterland T.,  
 RA Doolittle W.F., Ragan M.A., Charlebois R.L.,  
 RT "Organizational characteristics and information content of an  
 RT archaeal genome: 156 kb of sequence from Sulfolobus solfataricus  
 RT P2.",  
 RL Mol. Microbiol. 22:175-191(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=1142726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweyer M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moore A., Brauso G., Fletcher C., Gordon P.M.K.,  
 RA Hekamp-de Jong I., Jelfries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Thériault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sersen C.W., Van der Oost J.,  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 CC -1- CATALYTIC ACTIVITY: 4-carboxymethylglutamate-2-en-4-olide + H(2)O = 4  
 CC oxohex-2-enedioate.  
 CC -1- SIMILARITY: BELONGS TO THE DIENELACTONE HYDROLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y08256; CAA69498.1; -  
 DR EMBL: AE006815; AAK42256.1; -  
 DR InterPro: IPR002925; DLH.  
 DR Pfam: PF01738; DLH; 1.  
 KW Hypothetical protein; Hydrolase; Complete proteome.  
 FT ACT SITE 148 148 BY SIMILARITY.  
 FT ACT SITE 195 195 BY SIMILARITY.  
 FT ACT SITE 226 226 BY SIMILARITY.  
 SQ SEQUENCE 257 AA; 29016 MW; B2BD7FACD6126382 CRC64;  
 QY Query Match 16.9%; Score 56; DB 1; Length 257;  
 DB Local Similarity 32.2%; Pred. No. 5.7;  
 DB Local 19; Conservative 6; Mismatches 18; Indels 16; Gaps 3;  
 DB 129 KAYEVSSQGVKIVSMGFCWGG--LAFQ-----LATEVPIDGTIVFYGRNPP 176

RESULT 13  
 ID G55A CHICK STANDARD; PRT; 337 AA.  
 AC Q96892;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neurite inhibitor GP55-A precursor (OBSCAM protein gamma isoform).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OK NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Lodge A.P., Kim D.-S., Howard M.R., McNamee C.J., Smith N., Moss D.J.,  
 RT "Cloning of CEPU-S, a secreted isoform of CEPU-1, and OBSCAM cDNAs from  
 RT chick: structural diversity of IGLON family proteins.",  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]

SEQUENCE OF 64-337 FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Brain; PubMed=9004047;  
 RA TISSUE=Brain; PubMed=9715768;  
 RX Wilson D.J.A., Kim D.-S., Clarke G.A., Marshall-Clarke S., Moss D.J.,  
 "A family of glycoproteins (GP55), which inhibit neurite outgrowth,  
 are members of the Ig superfamily and are related to OSCM,  
 neurotrophin, LAMP and CEP-1."  
 RT J Cell Sci. 109:3129-3138(1996).  
 CC -1- FUNCTION: INHIBITS NEURITE OUTGROWTH.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: RESTRICTED TO THE NERVOUS SYSTEM.  
 CC -1- DEVELOPMENTAL STAGE: INCREASES DURING DEVELOPMENT FROM VERY LOW  
 CC LEVELS AT EMBRYONIC DAY 10 AND IS MOST ABUNDANT AFTER HATCHING.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON  
 CC SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Y08170; CAB41420.1; -  
 CC InterPro: IPR003006; IG\_MHC.  
 CC InterPro: IPR003598; IG\_C2.  
 CC InterPro: IPR003600; IG\_Like.  
 CC Pfam; PF00047; IG\_3.  
 CC SMART; SM00410; IG\_Like; 1.  
 CC DR Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 CC Repeat: Signal.

CC SIGNAL 1 20 BY SIMILARITY.  
 CC CHAIN 21 317 NEURITE INHIBITOR GP55-A (POTENTIAL).  
 CC PROPEP 318 337 REMOVED IN MATURE FORM (POTENTIAL).  
 CC DOMAIN 43 115 IG-LIKE C2-TYPE DOMAIN 1.  
 CC DOMAIN 143 201 IG-LIKE C2-TYPE DOMAIN 2.  
 CC DOMAIN 229 295 IG-LIKE C2-TYPE DOMAIN 3.  
 CC DISULFID 50 108 POTENTIAL.  
 CC DISULFID 150 194 POTENTIAL.  
 CC DISULFID 236 288 POTENTIAL.  
 CC CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC SEQUENCE 337 AA; 36887 MW; BAE71755185651E CRC64;

CC -1- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 15 BNR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB049473; BAB78470.1; -  
 CC InterPro: IPR000561; EGF-Like.  
 CC InterPro: IPR002860; GH\_BNR.  
 CC InterPro: IPR002861; Reeler.  
 CC Pfam; PF00008; EGF\_5.  
 CC Pfam; PF02012; BNR; 15.  
 CC Pfam; PF02014; Reeler; 1.  
 CC PROSITE; PS01186; EGF\_2; 6.  
 CC PROSITE; PS01186; EGF\_2; 6.  
 CC Hydroxylase; Serine protease; Developmental protein; Matrix protein;  
 CC Cell adhesion; EGF-like domain; Glycoprotein; Repeat; Signal;  
 CC Alternative splicing.

CC SIGNAL 1 27 POTENTIAL.  
 CC CHAIN 28 3462 REELIN.  
 CC DOMAIN 41 173 EGF-LIKE 1.  
 CC DOMAIN 672 703 EGF-LIKE 2.  
 CC DOMAIN 1031 1062 EGF-LIKE 3.  
 CC DOMAIN 1410 1443 EGF-LIKE 4.  
 CC DOMAIN 1766 1797 EGF-LIKE 5.  
 CC DOMAIN 2130 2162 EGF-LIKE 6.  
 CC DOMAIN 2479 2510 EGF-LIKE 7.  
 CC DOMAIN 2854 2885 EGF-LIKE 8.  
 CC DOMAIN 3229 3261 BNR 1.  
 CC REPEAT 594 605 BNR 2.  
 CC REPEAT 800 811 BNR 3.  
 CC REPEAT 953 964 BNR 4.  
 CC REPEAT 1158 1169 BNR 5.  
 CC REPEAT 1324 1335 BNR 6.  
 CC REPEAT 1536 1547 BNR 7.  
 CC REPEAT 1687 1698 BNR 7.

RC TISSUE=Cerebellum;  
 RA Kikawa S., Terashima T.;  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RX ALTERNATIVE SPLICING.  
 RP MEDLINE=99253436; PubMed=10328932;  
 RA Lambert de Rouvroit C., Bernier B., Royaux I., de Beyerck V.,  
 Goffinet A.M.;  
 RT "Evolutionarily conserved, alternative splicing of reelin during brain  
 development."  
 RL Exp. Neurol. 156:229-238(1999).  
 CC -1- FUNCTION: Extracellular matrix serine protease that plays a role  
 CC in layering of neurons in the cerebral cortex and cerebellum.  
 CC Regulates microtubule function in neurons and neuronal migration.  
 CC Affects migration of sympathetic preganglionic neurons in the  
 CC spinal cord, where it seems to act as a barrier to neuronal  
 CC migration. Enzymatic activity is important for the modulation of  
 CC cell adhesion. Binding to the extracellular domains of lipoprotein  
 CC receptors VLDLR and ApoER2 induces tyrosine phosphorylation of  
 CC Dab1 and modulation of Tau phosphorylation (By similarity).  
 CC -1- SUBUNIT: Binds to the ectodomains of VLDLR and ApoER2 (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; are  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Abundantly produced during brain ontogenesis  
 CC by the Cajal-Retzius cells and other pioneer neurons located in  
 CC the telencephalic marginal layer of the cerebellum.  
 CC external granular layer of the cerebellum.  
 CC -1- DOMAIN: The basic C-terminal region is essential for secretion (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE REELIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 15 BNR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB049473; BAB78470.1; -  
 CC InterPro: IPR000561; EGF-Like.  
 CC InterPro: IPR002860; GH\_BNR.  
 CC InterPro: IPR002861; Reeler.  
 CC Pfam; PF00008; EGF\_5.  
 CC Pfam; PF02012; BNR; 15.  
 CC Pfam; PF02014; Reeler; 1.  
 CC PROSITE; PS01186; EGF\_2; 6.  
 CC PROSITE; PS01186; EGF\_2; 6.  
 CC Hydroxylase; Serine protease; Developmental protein; Matrix protein;  
 CC Cell adhesion; EGF-like domain; Glycoprotein; Repeat; Signal;  
 CC Alternative splicing.

CC SIGNAL 1 27 POTENTIAL.  
 CC CHAIN 28 3462 REELIN.  
 CC DOMAIN 41 173 EGF-LIKE 1.  
 CC DOMAIN 672 703 EGF-LIKE 2.  
 CC DOMAIN 1031 1062 EGF-LIKE 3.  
 CC DOMAIN 1410 1443 EGF-LIKE 4.  
 CC DOMAIN 1766 1797 EGF-LIKE 5.  
 CC DOMAIN 2130 2162 EGF-LIKE 6.  
 CC DOMAIN 2479 2510 EGF-LIKE 7.  
 CC DOMAIN 2854 2885 EGF-LIKE 8.  
 CC DOMAIN 3229 3261 BNR 1.  
 CC REPEAT 594 605 BNR 2.  
 CC REPEAT 800 811 BNR 3.  
 CC REPEAT 953 964 BNR 4.  
 CC REPEAT 1158 1169 BNR 5.  
 CC REPEAT 1324 1335 BNR 6.  
 CC REPEAT 1536 1547 BNR 7.  
 CC REPEAT 1687 1698 BNR 7.

FT REPEAT 1885 1896 BNR 8.  
FT REPEAT 2044 2055 BNR 9.  
FT REPEAT 2251 2262 BNR 10.  
FT REPEAT 2400 2411 BNR 11.  
FT REPEAT 2599 2610 BNR 12.  
FT REPEAT 2779 2790 BNR 13.  
FT REPEAT 2980 2991 BNR 14.  
FT REPEAT 3364 3375 BNR 15.  
FT DOMAIN 3433 3462 ARG-RICH (BASIC).  
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1448 1448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1601 1601 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1922 1922 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2146 2146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2270 2270 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2318 2318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2570 2570 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2963 2963 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3017 3017 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3074 3074 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3186 3186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3413 3413 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3440 3440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 3430 3431 MISSING (IN ISOFORM 2).  
FT VARSPLIC 3430 3462 MISSING (IN ISOFORM 3).  
SQ SEQUENCE 3462 AA; 387525 MW; FCC898908035F6 CRC64;

Query Match 16.9%; Score 56; DB 1; Length 3462;  
Best Local Similarity 29.7%; Pred. No. 1e+02;  
Matches 11; Conservative 7; Mismatches 19; Indels 0; Gaps 0;  
QY 10 YVYIKNDGKKECHGAGNLEFGIRSEQINLSTRN 46  
Db 374 WLFPGATVYKHSQSDGNALYFPHGNSQLNPATTRD 410

RESULT 15  
MTPN CHICK STANDARD; PRT; 118 AA.  
AC 091955;  
DF 01-NOV-1997 (Rel. 35, Created)  
P 01-NOV-1997 (Rel. 35, Last sequence update)  
D 15-JUN-2002 (Rel. 41, Last annotation update)  
B Myotrophin (V-1 protein) (granule cell differentiation protein).  
GN MTPN.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A..  
RC Tissue: lens fibers;  
RX MEDLINE=96437509; PubMed=8840185;  
RA Sawada K., Agata K., Eguchi G.;  
RT "Characterization of terminally differentiated cell state by  
categorizing cDNA clones derived from chicken lens fibers.";  
RL Int. J. Dev. Biol. 40:531-535(1996)  
CC -!- FUNCTION: POTENTIAL ROLE IN CEREBELLAR MORPHOGENESIS. MAY FUNCTION  
IN DIFFERENTIATION OF CEREBELLAR NEURONS, PARTICULARLY OF GRANULE  
CELLS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: CONTAINS 3 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D26326; BAA05379.1; -  
DR HSSP; P80144; 2MVO.  
DR InterPro; IPR002110; ANK.  
DR Pfam; PF00023; ank; 3.  
DR SMART; SM00248; ANK; 2.  
DR PROSITE; PS00088; ANK\_REPEAT; 2.  
DR PROSITE; PS00297; ANK\_REGION; 1.  
KW Repeat; ANK repeat.  
FT REPEAT 1 30 ANK 1.  
FT REPEAT 34 65 ANK 2.  
FT REPEAT 67 98 ANK 3.  
SQ SEQUENCE 118 AA; 12886 MW; 6097EB55C6CBD2CF CRC64;

Query Match 16.7%; Score 55.5; DB 1; Length 118;  
Best Local Similarity 30.0%; Pred. No. 2.8;  
Matches 15; Conservative 8; Mismatches 22; Indels 5; Gaps 1;  
QY 16 DQKKKECHGAGN-----LLEFGIRSEQINLSTRNPCNITSRYVGHTS 60  
Db 33 EGGRKPLHYAADGQLEILFLLLKGNADINAPDKNITPLLSAVYBGHS 82

Search completed: March 10, 2003, 18:25:27  
Jed time : 1.98569 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:11:50 ; Search time 1.77424 Seconds

(without alignments)  
3359.364 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649

Sequence: 332 1 AANFIAGKRVYIKNDGKMK.....STRNPNITSRYVGGHTSPT 62

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Minimum DB seq length: 0 283224

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR 73:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	100.0	1196	2	T09356
2	128	38.6	1166	2	P96598
3	94.5	28.5	1143	2	B84431
4	63.5	19.1	1322	2	AD3265
5	63.5	19.1	1436	2	D71618
6	63	19.0	916	2	C75175
7	62	18.7	871	2	B71039
8	61	18.4	313	2	A64084
9	60	18.1	1227	2	AE1810
10	60	18.1	3078	2	T28432
11	59	17.8	323	2	H86716
12	58.5	17.6	316	1	S15799
13	58.5	17.6	776	2	C69072
14	58	17.5	342	2	S23764
15	58	17.5	397	2	U80082
16	57.5	17.3	523	2	G96777
17	57.5	17.3	728	1	HHCH90
18	57.5	17.3	732	1	HHH066
19	57.5	17.3	733	1	HHH066
20	57	17.2	107	2	G97835
21	57	17.2	300	2	B8327
22	57	17.2	300	2	AH2955
23	57	17.2	1488	2	C70984
24	56.5	17.0	907	2	T27317
25	56	16.9	257	2	S73082
26	55.5	16.7	180	2	H90340
27	55.5	16.7	180	2	H90340
28	55.5	16.7	656	1	X850KP
29	55.5	16.7	658	2	P95237

brassinosteroid-in protein F20N2.4 [1]  
probable receptor  
asparaginase (EC 3  
hypothetical prote  
hypothetical prote  
hypothetical prote  
lysophospholipase  
WD-40 repeat prote  
variant-specific s  
L-asparaginase [im  
protein-cytosine-P  
anaerobic ribonuci  
polygalacturamase-  
GPI-linked recepto  
hypothetical prote  
heat shock protein  
heat shock protein  
hypothetical prote  
hypothetical prote  
probable pps8 prot  
USF protein homolo  
hypothetical prote  
transketolase (EC  
transketolase [imp

30	55.5	16.7	658	2	G98101	transketolase (EC
31	55.5	16.7	1100	2	H71616	SERA antigen/papal
32	55.5	16.7	1192	2	F88062	protein ZK250.10 [
33	55.5	16.7	1360	2	T32833	hypothetical prote
34	55.5	16.7	1466	2	T32422	hypothetical prote
35	55	16.6	322	2	G89926	hypothetical prote
36	55	16.6	359	2	G95955	conserved probable
37	55	16.6	445	2	T20190	hypothetical prote
38	55	16.6	467	2	A57627	p55 erythrocyte me
39	55	16.6	474	2	E64207	hypothetical prote
40	55	16.6	890	2	A48753	NFAT transcription
41	54.5	16.4	384	2	F32252	gene III protein-
42	54.5	16.4	414	2	T23940	hypothetical prote
43	54.5	16.4	573	2	A41319	3-oxosteroid 1-deh
44	54	16.3	155	2	C97052	choistamate mutase
45	54	16.3	211	2	B84066	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T09356

brassinosteroid-insensitive protein BR1 - Arabidopsis thaliana

N/Alternate names: protein F23K16.30

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 24-Nov-1999

C/Accession: T09356

R/By: M. J. Murphy, G. J. Ridley, P. J. Hudson, S. J. Bancroft, I. J. Mewes, H. W. J. Mayer, K. F. X.

Submitted to the Protein Sequence Database, June 1999

A/Reference number: Z16552

A/Accession: T09356

A/Molecule type: DNA

A/Residues: 1-1196 <BEV>

A/Cross-References: EMBL:AU078620; GSPDB:GN00062; ATSP:F23K16.30

C/Experimental source: cultivar Columbia; BAC clone F23K16

C/Genetics:

A/Map position: 4

C/Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; p

Query Match

Best Local Similarity 100.0%; Score 332; DB 2; Length 1196;

Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 AANFIAGKRVYIKNDGKMKCHGAGNLEFGQIRSBQNLNSTRNPNITSRYVGGHTS 60

Db

588 AANFIAGKRVYIKNDGKMKCHGAGNLEFGQIRSBQNLNSTRNPNITSRYVGGHTS 647

Qy

61 PT 62

Db

648 PT 649

RESULT 2

P96598

protein F20N2.4 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: P96598

R/Theologos, A.; Ecker, J. R.; Palm, C. J.; Federspiel, N. A.; Kaul, S.; White, O.; Alonso,

Chen, C. W.; Chung, M. K.; Conn, L. J.; Conway, A. B.; Conway, A. R.; Creasy, T. H.; Dewar, K.;

ansen, N. F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J. L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.

C. A.; Li, J. H.; Li, Y.; Lin, X.; Liu, S. X.; Liu, Z. A.; Luros, J. S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S. L.; Schwartz, J. R.; Shinn, P.; Southwick, A. M.; Sun, H.; Tallon, I.

ker, M.; Wu, D. J.; Yu, G.; Fraser, C. M.; Venter, J. C.; Davis, R. W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A66141; MUID:21016719; PMID:11130712  
A/Accession: P96598  
A/Status: preliminary



M.; Ohnuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998  
 A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic A/Reference number: A71000; MUID:98344137; PMID:9679194  
 A/Accession: B71039  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-871 <K&M>  
 A/Cross-references: GB:AP00006; NID:93336133; PIDN:BA30714.1; PID:93258031  
 A/Experimental source: strain OT3  
 A/Note: this accession replaces an interim accession for a sequence replaced by GenBank C/Genetics:  
 A/Gene: PH1602

Query Match 18.7%; Score 62; DB 2; Length 871;  
 Best Local Similarity 37.0%; Pred. No. 14;  
 Matches 17; Conservative 6; Mismatches 9; Indels 14; Gaps 3;  
 Oy 10 YVYIKNDGKKE-----CHGAGNLT-----EFG--IRSEQLNR 41  
 D 776 YILAGTGAKEITFGSCGAGNLSRKATRYQYRGRIRQLNLR 821

RESULT 8  
 A64084  
 Iysohophilipase I2 homolog - Haemophilus influenzae (strain Rd KW20)  
 C/Species: Haemophilus influenzae  
 C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
 C/Accession: A64084  
 R/Plaschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirschner, E.F.; Kerlavage, J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995  
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A/Reference number: A64080; MUID:95350630; PMID:7542800  
 A/Accession: A64084  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-313 <TIGR>  
 A/Cross-references: GB:U32747; GB:I42023; NID:91573635; PIDN:AC22305.1; PID:91573643; T C/Suprafamily: Iysohophilipase I2

Query Match 18.4%; Score 61; DB 2; Length 313;  
 Best Local Similarity 36.6%; Pred. No. 6;  
 Matches 15; Conservative 9; Mismatches 15; Indels 2; Gaps 1;  
 Oy 7 GRRVYIKNDGKKECHGAGNLTFFQIRSEQLNRSTRNP 47  
 D 181 GRRVYFGK--GAYQQAHDYVELTFCRTRMKWNRINRKNP 219

RESULT 9  
 AE1810  
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
 C/Species: Nostoc sp.  
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C/Accession: AE1810  
 R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuytaz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, DNA Res. 9, 205-213, 2001  
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: AE1810  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1227 <KUR>  
 A/Cross-references: GB:BA000019; PIDN:BA077553.1; PID:917135007; GSPDB:GN00179  
 A/Experimental source: strain PCC 7120  
 C/Genetics:  
 A/Gene: alr0029

Query Match 18.1%; Score 60; DB 2; Length 1227;  
 Best Local Similarity 25.5%; Pred. No. 36;  
 Matches 12; Conservative 16; Mismatches 17; Indels 2; Gaps 1;  
 Oy 13 IKNDGKKECHGAGNLTFFQIRSEQLNRSTRNPNTSRVYGH 59  
 Db 528 LQKEPLPKGYAAGNLTIN--LRQLQDKTPEPSFIDSGRDFSLT 572

RESULT 10  
 T28432  
 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)  
 N/Alternate names: erythrocyte membrane binding protein 1 (EMP1)  
 C/Species: Plasmodium falciparum  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
 C/Accession: T28432  
 R/Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfeldt, J.A.; Peterson, D.S. Cell 82, 89-100, 1995  
 A/Title: The large diverse gene family var encodes proteins involved in cytoadherence and A/Reference number: Z20487; MUID:95330813; PMID:7606788  
 A/Accession: T28432  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-3078 <STX>  
 A/Cross-references: EMBL:U40608; NID:9886374; PID:9886375; PIDN:AAA75396.1  
 C/Genetics:  
 A/Gene: var-1  
 A/Introns: 2611/3

Query Match 18.1%; Score 60; DB 2; Length 3078;  
 Best Local Similarity 31.4%; Pred. No. 99;  
 Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;  
 Oy 16 DQMKKECHGAGNLTFFQIRSEQLNRSTRNPNC-----NITSRYVGH 58  
 Db 2278 DELENACRGAG---IFEGIRDE--WKCRVCGYVCKPENVNGEAKGH 2322

RESULT 11  
 H86716  
 L-asparaginase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C/Species: Lactococcus lactis subsp. lactis  
 C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C/Accession: H86716  
 R/Bolotin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich Genome Res. 11, 731-753, 2001  
 A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp A/Reference number: A86625; MUID:21235186; PMID:11337471  
 A/Accession: H86716  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-323 <STO>  
 A/Cross-references: GB:AE005176; PID:912723651; PIDN:AAK04834.1; GSPDB:GN00146  
 A/Experimental source: strain IL1403  
 C/Genetics:  
 A/Gene: amB  
 C/Suprafamily: asparaginase

Query Match 17.8%; Score 59; DB 2; Length 323;  
 Best Local Similarity 38.3%; Pred. No. 11;  
 Matches 18; Conservative 5; Mismatches 20; Indels 4; Gaps 2;  
 Oy 16 DQMKKECHGAGNLTFFQIRSEQLNR-LSTRNPNTSRVYGH 61  
 Db 229 DGVIVTALGAGNT---PIASGLQLRLDAKVPVVLVSRCFNGIAEP 272

RESULT 12  
 S15799  
 protein-tyrosine-phosphatase (EC 3.1.3.48) cdcs2 homolog - Caenorhabditis elegans  
 N/Alternate names: protein ZK637.11  
 C/Species: Caenorhabditis elegans  
 C/Date: 06-Jan-1995 #sequence\_revision 25-Apr-1997 #text\_change 20-Mar-1998

A/Accession: S15799  
R/Craxon, M.; Ainscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; Hall,  
Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterson, R.; Wilson, R.  
Submitted to the EMBL Data Library, May 1991

A/Reference number: S15786  
A/Acession: S15799  
A/Molecule type: DNA  
A/Residues: 1-316 <CR>  
A/Cross-references: EMBL:Z11115; NID:g96953; PID:g6964

C/Genetics:  
A/Functon: 34/2, 157/3  
C/Description: catalyzes hydrolysis of peptidyl-phosphotyrosine to peptidyl-L-tyrosine and  
A/Pathway: initiation of mitosis  
A/Note: cdc25 activates the cdc2 protein kinase by dephosphorylating it  
C/Superfamily: Caenothobdittis protein-tyrosine-phosphatase cdc25 homolog; cdc25-type pro  
C/Keywords: phosphoprotein; phosphorinosomester hydrolase; tyrosine-specific phosphatae  
F/R/2/74/Domin: cdc25-type protein-tyrosine-phosphatase homology <PTP>  
F/189/Active site: Cys (phosphocysteine intermediate) #status predicted  
F/Binding site: substrate phosphate (Arg) #status predicted

Query Match 17.6%; Score 58.5; DB 1; Length 316;  
Best Local Similarity 27.3%; Pred. No. 13;  
Matches 15; Conservative 13; Mismatches 22; Indels 5; Gaps 2;

RESULT 13  
065072 anaerobic ribonucleoside-triphosphate reductase - Methanobacterium thermoautotrophicum  
C/Species: Methanobacterium thermoautotrophicum  
C/Pdate: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C/Smith, D.R.; Doucet-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Qiu, D.; Spadafora, R.; Vicalre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Giovanni, N.  
J.; Church, G.M.; Daniels, C.U.; Mao, J.; Rice, P.; Noelting, T.P.; Reeve, J.N.

F/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
A/Reference number: A69000; PMID:98037514; PMID:9377463

A/Accession: C69072  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-776 <MTX>  
A/Cross-references: GB:AEO00914; GB:AEO00666; NID:g2622656; PIDN:AB86013.1; PID:g262265  
A/Experimental source: strain Delta H  
A/Notes:

A/Name: MTM1539  
A/Start codon: TTG

Query Match 17.6%; Score 58.5; DB 2; Length 776;  
Best Local Similarity 36.4%; Pred. No. 34;  
Matches 16; Conservative 6; Mismatches 19; Indels 3; Gaps 1;

5 TAGKRVYIKNDGMKKCKAGCNILFEQ--GIHQEQLNLSTR 45  
| | : | | | | | | : | | | : | | | |  
Dn 541 IDGRYRIENATWTSFGFTGNEMLEHLAGTIGSPANFRGLR 584

RESULT 14  
S23764 polygalacturonase-inhibiting protein precursor - kidney bean  
C/Species: Phaseolus vulgaris (kidney bean)  
C/Pdate: 05-Mar-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C/Accession: S23764  
R/Tombart, P.; Desiderio, A.; Salvi, G.; Cervone, F.; Daroda, L.; de Lorenzo, G.; Bergma  
Plant J., 2, 367-373, 1992  
A/Title: Cloning and characterization of the gene encoding the endopolygalacturonase-inh  
A/Reference number: S23764; PMID:93272053; PMID:1103801  
A/Accession: S23764  
A/Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-342 <NO>
A:Cross-references: EMBL:X64769; NID:g21028; PIDs:CA46016.1; PID:g21029
C>Note: It is uncertain whether Met-1 or Met-10 is the initiator
C:superfamily: polygalacturanase-inhibiting protein; leucine-rich alpha-2-glycoprotein repeat homology <LRR>
F:274-297/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>

Query Match          17.5%; Score 58; DB 2; Length 342;
Best Local Similarity 34.5%; Pred. No. 16;
Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;

QY      3   NFPIAGKRVYVTKNDGMKEKGAGNLTFFQ----GIRSEQLNLSTENPCNITSRYVG 56
       :|::||::||:|||||:|||||:|::|||:|::|||:|::|||
Db       236  NWLEGDASVLFGSDKNTRYKTHLAQNLSLAFDLGKGL-SKNLNGLDLRN-----NRIYG 287

RESULT 15
JR0082
GPI-linked receptor precursor - mouse
M:Alternate names: GFalpha-3
C:Species: Mus musculus (house mouse)
C:Date: 21-May-1998 #sequence_revision 29-May-1998 #text_change 02-Jun-2000
C:Accession: JR0082
R:Nomoto, S.; Ito, S.; Yang, L.X.; Kuchi, K.
Biochem. Biophys. Res. Commun. 244, 849-853, 1998
A>Title: Molecular cloning and expression analysis of GFalpha-3, a novel CDNA related to
A:reference number: JF0082; MUID:98205811; PMID:9535755
A:Accession: JB0082
A:Molecule type: mRNA
A:Residues: 1-397 <NO>
A:Cross-references: DDBJ:AB008833; NID:g2627159; PIDs:BAA23552.1; PID:g2627160
C:Comment: This protein plays a distinct role in cell survival and differentiation.
C:superfamily: Mus musculus GPI-linked receptor
C:keywods: glycoprotein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:380-397/Region: hydrophobic
F:92,145,306/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match          17.5%; Score 58; DB 2; Length 397;
Best Local Similarity 44.4%; Pred. No. 19;
Matches 12; Conservative 4; Mismatches 7; Indels 4; Gaps 1;

QY      22  CHQGNLTLPQGIKRSQNLSTRNPC 48
       :|::||:|||||:|::|||:|::|||
Db       315  CRGSQNLQD---ECEQLERSFSQNPC 337

Search completed: March 10, 2003, 18:24:54
Job time : 3.77424 secs
```

```

; GENERAL INFORMATION:
; APPLICANT: Collier, Alan
; APPLICANT: Alfano, James R

```

APPLICANT: Charkowski, Amy O.  
TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE  
TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES  
FILE REFERENCE: 19603/3243  
CURRENT APPLICATION NUMBER: US/09/825,414  
CURRENT FILING DATE: 2001-04-03  
PRIOR APPLICATION NUMBER: 60/194,160  
PRIOR FILING DATE: 2000-04-03  
PRIOR APPLICATION NUMBER: 60/224,604  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/249,548  
PRIOR FILING DATE: 2000-11-17  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 466  
TYPE: PRT  
ORGANISM: Pseudomonas syringae  
US-09-825-414-20

Query Match  
Best Local Similarity 18.2%; Score 60.5; DB 10; Length 466;  
St Local Similarity 26.4%; Pred. No. 4.3;  
Matches 19; Conservative 11; Mismatches 31; Indels 11; Gaps 2;

Qy 1 AANFAGKRYVYIKNDGM-----KKECHGAGNLFEGGIRSEQLN--RLSTRNPN 49  
Db 387 AMNYVAAKIRLSKPEGKVAFAVGATHTATSCDVPGLAEHLHGVSRLVLDGLKSRATVD 446

Qy 50 ITRVYGGHTSP 61  
Db 447 INVKNYGGKLNK 458

RESULT 3  
US-10-108-605-127  
Sequence 127; Application US/10108605  
Patent No. US20020160934A1  
GENERAL INFORMATION:  
APPLICANT: Broadus, Julie  
APPLICANT: Stam, Lynn  
APPLICANT: Bachmann, Jane  
APPLICANT: Kamdar, Kim  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE  
TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF  
FILE REFERENCE: 31133B  
CURRENT APPLICATION NUMBER: US/10/108,605  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: US 09/761,142  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/176,418  
PRIOR FILING DATE: 2000-01-14  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 127  
LENGTH: 808  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-108-605-127

Query Match  
Best Local Similarity 18.2%; Score 60.5; DB 9; Length 808;  
St Local Similarity 38.0%; Pred. No. 8.4;  
Matches 19; Conservative 5; Mismatches 23; Indels 3; Gaps 1;

Qy 2 ANFIAGKRYVYIKNDGMKKECHGAGNLFEGGIRSEQLNRLSTRNPN 51  
Db 727 ARFKQGRVYVYEDNGINP---GAFNPLQIQIRKVTARLLCDNDRLL 773

RESULT 4  
US-10-153-273-12  
Sequence 12; Application US/10153273  
Patent No. US20020169305A1  
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
Chitnis, Chetan  
Miller, Louis H.  
Peterson, David S.  
Su, Xin-zhaun  
Wellems, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/153,273  
FILING DATE: 21-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,288  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Fuller, Michael  
REGISTRATION NUMBER: 36,516  
REFERENCE/DOCKET NUMBER: NIH121.1PMDV1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDNESS: single  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-10-153-273-12

Query Match  
Best Local Similarity 18.1%; Score 60; DB 9; Length 2710;  
St Local Similarity 31.4%; Pred. No. 43;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

Qy 16 DNMKKECHGAGNLFEGGIRSEQLNRLSTRNPN-----NITSRVYGGH 58  
Db 2280 DLENACRGAG---IFEGIRKDE---WKCRNVCGYVVCPEPNNNGAKKH 2324

RESULT 5  
US-09-220-920-64  
Sequence 64; Application US/09220920  
Patent No. US2002002269A1  
GENERAL INFORMATION:  
APPLICANT: Milbrandt, Jeffrey D.  
APPLICANT: Balogh, Robert H.  
TITLE OF INVENTION: Artemin, A No. US2002002269A1 Neurotrophic Factor  
FILE REFERENCE: 6029-7996  
CURRENT APPLICATION NUMBER: US/09/220,920  
CURRENT FILING DATE: 1998-12-24  
EARLIER APPLICATION NUMBER: 09/163,283  
EARLIER FILING DATE: 1998-09-29  
EARLIER APPLICATION NUMBER: 60/108,148  
EARLIER FILING DATE: 1998-11-12  
EARLIER APPLICATION NUMBER: 09/218,698



EARLIER FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 64  
LENGTH: 397  
TYPE: PRT  
ORGANISM: Murine  
US-09-220-920-64

Query Match 17.5%; Score 58; DB 10; Length 397;  
Best Local Similarity 44.4%; Pred. No. 7.6;  
Matches 12; Conservative 4; Mismatches 7; Indels 4; Gaps 1;

QY 22 CHGAGNLEFGGIRSEQLNR--LSTRNPNITSRVYG 48  
DB 315 CRGSGNLD---RCEQLERFSQNPC 337

RESULT 6  
US-09-496-18  
Sequence 18, Application US/09991496  
Patent No. US20020169285A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Melo, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
FILE REFERENCE: 210121.420C9  
CURRENT APPLICATION NUMBER: US/09/991,496  
CURRENT FILING DATE: 2001-11-20  
NUMBER OF SEQ ID NOS: 137  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-991-496-18

Query Match 17.3%; Score 57.5; DB 9; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
QY 14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNITSRVYG 56  
DB 269 KDDGDKK---KKIKKEYIDGELNKTPIWTRNPDITNEYG 310

RESULT 7  
US-09-932-257A-25  
Sequence 25, Application US/09932257A  
Publication No. US20030039658A1  
GENERAL INFORMATION:  
APPLICANT: Estable, Mario  
APPLICANT: Koeder, Robert  
TITLE OF INVENTION: MCEP, A No. US20030039658A1 Transcription Factor  
FILE REFERENCE: 600-1-269N  
CURRENT APPLICATION NUMBER: US/09/932,257A  
CURRENT FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/226,340  
PRIOR FILING DATE: 2000-08-18  
PRIOR APPLICATION NUMBER: 60/226,339  
PRIOR FILING DATE: 2000-08-18  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 25  
LENGTH: 732

TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-932-257A-25

Query Match 17.3%; Score 57.5; DB 9; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNITSRVYG 56  
DB 269 KDDGDKK---KKIKKEYIDGELNKTPIWTRNPDITNEYG 310

RESULT 8  
US-09-759-010-5  
Sequence 5, Application US/09759010  
Patent No. US20010034042A1  
GENERAL INFORMATION:  
APPLICANT: Srivastava, Pramod K.  
TITLE OF INVENTION: COMPLEXES OF PEPTIDE BINDING FRAGMENTS OF HEAT-SHOCK  
TITLE OF INVENTION: PROTEINS AND THEIR USE AS IMMUNOTHERAPEUTIC AGENTS  
FILE REFERENCE: 8449-135  
CURRENT APPLICATION NUMBER: US/09/759,010  
CURRENT FILING DATE: 2001-01-12  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 5  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-759-010-5

Query Match 17.3%; Score 57.5; DB 10; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNITSRVYG 56  
DB 269 KDDGDKK---KKIKKEYIDGELNKTPIWTRNPDITNEYG 310

RESULT 9  
US-09-874-923-18  
Sequence 18, Application US/09874923  
Patent No. US20020081320A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Melo, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
FILE REFERENCE: 210121.420C8  
CURRENT APPLICATION NUMBER: US/09/874,923  
CURRENT FILING DATE: 2001-06-04  
NUMBER OF SEQ ID NOS: 122  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-874-923-18

Query Match 17.3%; Score 57.5; DB 10; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNITSRVYG 56

Db 269 KKGDKKK-----KKKIKKKYIDKEELNKTKPIWTRNPDITNEYG 310

## RESULT 10

US-10-007-262-1  
Sequence 1, Application US/10007262  
Patent No. US20020164748A1  
GENERAL INFORMATION:  
APPLICANT: Bistrup, Annette  
APPLICANT: Rosen, Steven D.  
APPLICANT: Tangemann, Kirsten  
APPLICANT: Hemmerich, Stefan  
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
FILE REFERENCE: 6510-107CIP  
CURRENT APPLICATION NUMBER: US/10/007,262  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 3.0  
ID NO 1  
LENGTH: 386  
TYPE: PRT  
ORGANISM: H. sapiens  
US-10-007-262-1

## Query Match

Best Local Similarity 17.2%; Score 57; DB 9; Length 386;  
Pred. No. 9.9;  
Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 12 YIKNDGKKECHGAGNLLLEFGIRSEQ-----LNRLST 44  
Db 340 YEKVSRLOKACGDAMNLLGYRHVRSSEQQRNLLDLST 378

## RESULT 11

US-09-816-825-2  
Sequence 2, Application US/09816825  
Patent No. US20010051370A1  
GENERAL INFORMATION:  
APPLICANT: Bistrup, Annette  
APPLICANT: Rosen, Steven D.  
APPLICANT: Hemmerich, Stefan  
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
FILE REFERENCE: 6510-107CON  
CURRENT APPLICATION NUMBER: US/09/816,825  
CURRENT FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 09/045,284  
PRIOR FILING DATE: 1998-03-20  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FastSeq for Windows Version 4.0  
ID NO 2  
LENGTH: 386  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-816-825-2

## Query Match

Best Local Similarity 17.2%; Score 57; DB 10; Length 386;  
Pred. No. 9.9;  
Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 12 YIKNDGKKECHGAGNLLLEFGIRSEQ-----LNRLST 44  
Db 340 YEKVSRLOKACGDAMNLLGYRHVRSSEQQRNLLDLST 378

## RESULT 12

US-09-988-982-3  
Sequence 3, Application US/09988982  
Patent No. US20020081699A1  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.

Shah, Puri  
Murry, Lynn E.

TITLE OF INVENTION: NOVEL HUMAN LYSOPHOSPHOLIPASE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/988,982

FILING DATE: 19-No. US20020081699A1-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/213,394

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0269 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 230 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 552244

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Query Match 16.4%; Score 54.5; DB 10; Length 230;  
Best Local Similarity 37.8%; Pred. No. 11;  
Matches 17; Conservative 7; Mismatches 14; Indels 7; Gaps 3;

QY 21 ECHGAGN---LLEFGIRSEQLNRLSTRPCNITSRVYG--HTS 60  
Db 168 QCHGDCDPLVPLMFGSLYVERLKLGV--NPANVTKVTYEGMMHSS 210

## RESULT 13

US-09-796-692-1724  
Sequence 1724, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mamion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FaSeq for Windows Version 3.0  
SEQ ID NO 1724  
LENGTH: 62  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-761-48451

Query Match 15.7%; Score 52; DB 9; Length 62;  
Best Local Similarity 44.8%; Pred. No. 4.9;  
Matches 13; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

29 LEEGIRSEQLNLSIRPNITSRYVG 57  
19 LTTGIRFEDNGIYFCQKCNMTSEVYOG 47

RESULT 14

US-09-864-761-48451  
Sequence 48451, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 48451  
LENGTH: 63  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004859.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.77  
OTHER INFORMATION: EST HUMAN HIT: AA077526.1, EVALU 8.00e-09  
OTHER INFORMATION: SWISSPROT HIT: Q13751, EVALU 4.40e+00  
US-09-864-761-48451

Query Match 15.7%; Score 52; DB 10; Length 63;  
Best Local Similarity 31.2%; Pred. No. 5;  
Matches 10; Conservative 8; Mismatches 12; Indels 2; Gaps 1;

10 YYIINDGKKECH--GAGNLLFEQIRSEQL 39  
19 YSLVGCGAGHQRCHVEVGVLDQDTGLVSRV 50

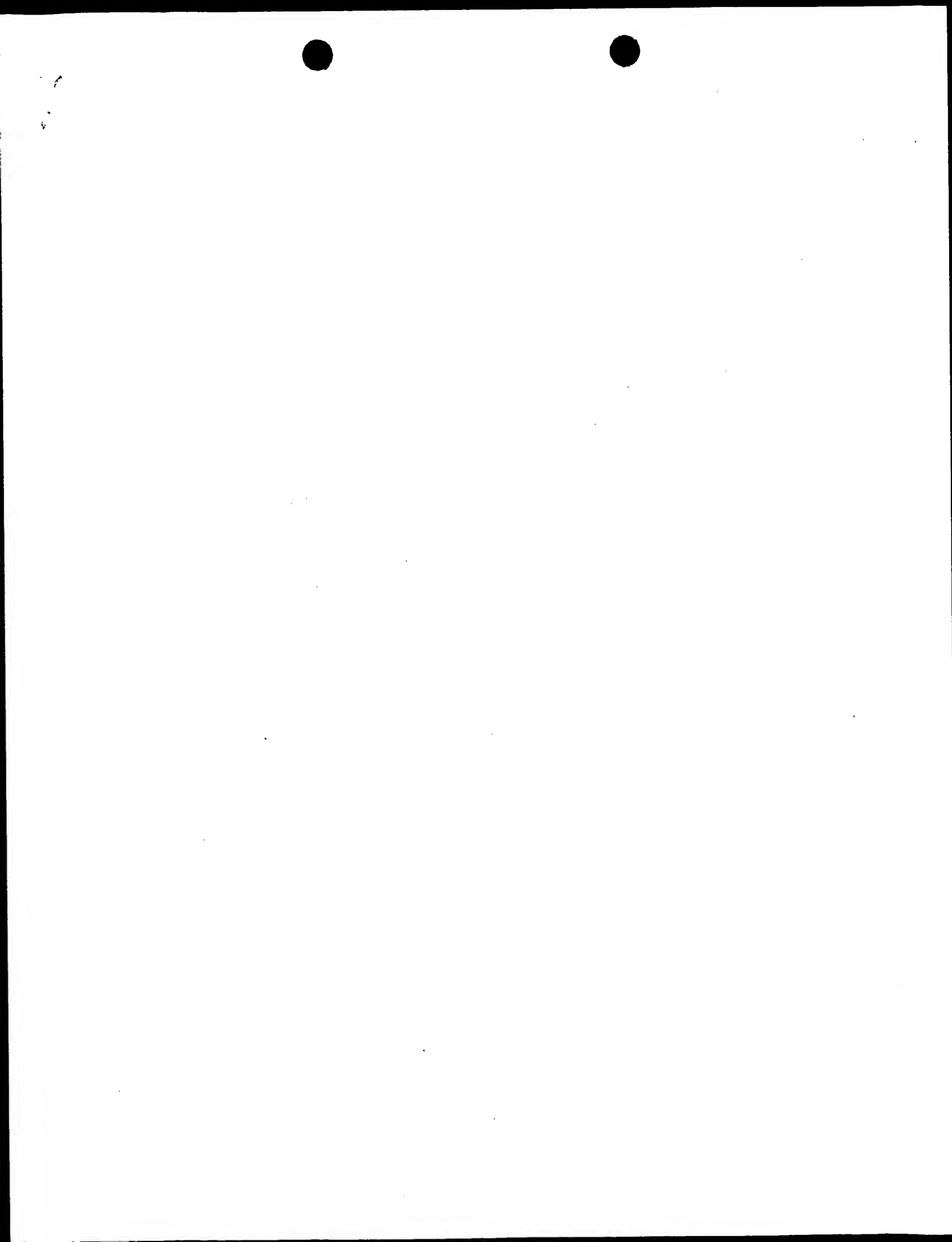
RESULT 15

US-10-038-107A-3  
Sequence 3, Application US/10038107A  
Patent No. US20020150573A1  
GENERAL INFORMATION:  
APPLICANT: Nusse, Michel  
TITLE OF INVENTION: ANTI-19 ALPHA-BETA ANTIBODY FOR LYMPHOMA THERAPY  
FILE REFERENCE: 7529/0405  
CURRENT APPLICATION NUMBER: US/10/038,107A  
CURRENT FILING DATE: 2002-04-17  
PRIOR APPLICATION NUMBER: 60/247,079  
PRIOR FILING DATE: 2000-10-11  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-107A-3

Query Match 15.7%; Score 52; DB 12; Length 229;  
Best Local Similarity 44.8%; Pred. No. 24;  
Matches 13; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

29 LEEGIRSEQLNLSIRPNITSRYVG 57  
107 LTTGIRFEDNGIYFCQKCNMTSEVYOG 135

Search completed: March 10, 2003, 18:11:43  
Job time: 2.33068 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:03:39 ; Search time 1.33068 Seconds  
(Without alignments)  
1370.890 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649  
Sequence: 332  
1 AANFIAGKRYVYIKNDGKMK.....STRNPNITSRVYGHTSPT 62

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

1: Issued Patents AA:\*  
2: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/6D\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	100.0	1196	4	US-08-881-706-2
2	60	18.1	2710	2	US-08-568-459A-12
3	60	18.1	2710	2	US-08-487-826B-12
4	60	18.1	2710	4	US-09-210-288-12
5	60	18.1	3060	2	US-08-487-826B-14
6	58	17.5	342	1	US-08-244-646-15
7	58	17.5	342	1	US-08-592-936B-21
8	58	17.5	342	2	US-09-111-573-21
9	58	17.5	397	4	US-09-220-528-64
10	57.5	17.3	732	2	US-08-533-669A-18
11	57.5	17.3	732	4	US-09-307-143-4
12	57.5	17.3	732	4	US-09-183-861-18
13	57.5	17.3	732	4	US-09-022-765-18
14	57.5	17.3	732	4	US-09-045-284A-2
15	57	17.2	386	4	US-09-190-911-1
16	55	16.6	130	2	US-08-467-046-12
17	55	16.6	890	1	US-08-145-006C-5
18	55	16.6	890	5	PCT-US94-00545-5
19	54.5	16.4	230	2	US-08-844-120-3
20	54.5	16.4	230	3	US-09-022-940-5
21	54.5	16.4	230	3	US-09-216-001-3
22	54.5	16.4	230	3	US-09-216-386-5
23	54.5	16.4	230	4	US-08-878-862-3
24	54.5	16.4	230	4	US-09-213-394-3
25	54.5	16.4	230	4	US-08-200-512-2
26	54	16.3	227	1	US-08-244-646-17
27	53	16.0	432	2	US-08-700-152A-4

28	53	16.0	525	1	US-08-077-939-19	Sequence 19, Appl
29	53	16.0	525	1	US-08-461-598-19	Sequence 19, Appl
30	53	16.0	525	1	US-08-461-621-19	Sequence 19, Appl
31	53	16.0	525	1	US-08-465-334-19	Sequence 19, Appl
32	52	15.7	476	1	US-08-313-075A-30	Sequence 30, Appl
33	52	15.7	553	4	US-08-997-251-2	Sequence 2, Appl
34	51.5	15.5	207	2	US-09-022-940-1	Sequence 1, Appl
35	51.5	15.5	207	2	US-09-022-940-1	Sequence 1, Appl
36	51.5	15.5	208	2	US-08-844-120-1	Sequence 1, Appl
37	51.5	15.5	208	2	US-09-213-394-1	Sequence 1, Appl
38	51.5	15.5	230	4	US-09-022-940-3	Sequence 3, Appl
39	51.5	15.5	230	3	US-09-216-386-3	Sequence 3, Appl
40	51	15.4	652	4	US-09-134-001C-3517	Sequence 3517, Ap
41	51	15.4	906	1	US-08-220-151-9	Sequence 9, Appl
42	51	15.4	906	1	US-08-413-118-9	Sequence 9, Appl
43	51	15.4	906	3	US-08-473-446-9	Sequence 9, Appl
44	50.5	15.2	118	3	US-08-965-904B-2	Sequence 2, Appl
45	50.5	15.2	118	4	US-08-934-131-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-881-706-2  
; Sequence 2, Application US/08881706  
; Patent No. 6245969  
; GENERAL INFORMATION:  
; APPLICANT: Chong, Joane  
; APPLICANT: Li, Jianming  
; TITLE OF INVENTION: Receptor Kinase BIN1  
; FILE REFERENCE: 07251/022001  
; CURRENT APPLICATION NUMBER: US/08/881,706  
; CURRENT FILING DATE: 1997-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1196  
; TYPE: PRT  
; ORGANISM: Arabidopsis sp.  
US-08-881-706-2

Query Match 100.0%; Score 332; DB 4; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 3.6e-38;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AANFIAGKRYVYIKNDGKMKCHGAGNLEFGIRSEQNLRLSTRNPNITSRVYGHTS 60  
DB 588 AANFIAGKRYVYIKNDGKMKCHGAGNLEFGIRSEQNLRLSTRNPNITSRVYGHTS 647  
QY 61 PT 62  
DB 648 PT 649

RESULT 2  
US-08-568-459A-12  
; Sequence 12, Application US/08568459A  
; Patent No. 5849306  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach

STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelisen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-568-459A-12

Query Match 18.1%; Score 60; DB 2; Length 2710;  
Best Local Similarity 31.4%; Pred. No. 32;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

QY 16 DGMKECHGAGNLEFQIGRSEQLRSTRNPC-----NITSRYVGGH 58  
DB 2280 DLENAACRGAG--IFSGIRKDE--MKCRNVCGYVCKPENVNGEAKGKH 2324

RESULT 3  
US-08-487-826B-12  
Sequence 12, Application US/08487826B  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelisen, Ned  
REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-487-826B-12

Query Match 18.1%; Score 60; DB 2; Length 2710;  
Best Local Similarity 31.4%; Pred. No. 32;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

QY 16 DGMKECHGAGNLEFQIGRSEQLRSTRNPC-----NITSRYVGGH 58  
DB 2280 DLENAACRGAG--IFSGIRKDE--MKCRNVCGYVCKPENVNGEAKGKH 2324

RESULT 4  
US-09-210-288-12  
Sequence 12, Application US/09210288  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,288  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fuller, Michael  
REGISTRATION NUMBER: 36,516  
REFERENCE/DOCKET NUMBER: NIH121.1FWDVL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-09-210-288-12

Query Match 18.1%; Score 60; DB 4; Length 2710;  
Best Local Similarity 31.4%; Pred. No. 32;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

QY 16 DQMKKECHGAGNLEFGIRSEQLNRLSTRNPC-----NITSRYVGGH 58  
DB 2280 DGLNACRGAG---IFSGIRKDE---WKCRNVCGYVCKPEVNVGEAKGKH 2324

RESULT 5  
US-08-487-826B-14

Sequence 14, Application US/08487826B  
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbie Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487, 826B

FILING DATE: 10-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelisen, Ned

REGISTRATION NUMBER: 29, 655

REFERENCE/DOCKET NUMBER: NIH121.001CP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 3060 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-487-826B-14

Query Match 18.1%; Score 60; DB 2; Length 3060;  
Best Local Similarity 31.4%; Pred. No. 37;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

QY 16 DQMKKECHGAGNLEFGIRSEQLNRLSTRNPC-----NITSRYVGGH 58  
DB 2278 DGLNACRGAG---IFSGIRKDE---WKCRNVCGYVCKPEVNVGEAKGKH 2322

RESULT 6  
US-08-244-646-15

Sequence 15, Application US/08244646  
Patent No. 5744692

GENERAL INFORMATION:

APPLICANT: Cervone, Felice

APPLICANT: De Lorenzo, Giulia

APPLICANT: Salvi, Giovanni

APPLICANT: Albersheim, Peter

APPLICANT: Davy, Alan

APPLICANT: Bergmann, Carl

TITLE OF INVENTION: Nucleotide Sequences Coding An

TITLE OF INVENTION: Endopolygalacturonase Inhibitor

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sally A. Sullivan

STREET: 5370 Manhattan Circle Suite 201

CITY: Boulder

STATE: CO

COUNTRY: US

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/244, 646

FILING DATE: 06-JUN-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: IT RM 91A 000915

FILING DATE: 06-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/IT/00158

FILING DATE: 04-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Sullivan, Sally A.

REGISTRATION NUMBER: 32,064

REFERENCE/DOCKET NUMBER: 19-94

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 342 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-244-646-15

Query Match 17.5%; Score 58; DB 1; Length 342;  
Best Local Similarity 34.5%; Pred. No. 4.3;  
Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;

QY 3 NFVAGKRYVYIKDKMKKECHGAGNLEFG---GIRSEQLNRLSTRNPCNITSRYVG 56  
DB 236 NMLEGDASVLFSGDKNTKTKIHLAKNSLAFLGKVGVL-SKVLNGLDLRN-----NRIYV 287

RESULT 7  
US-08-592-936B-21

Sequence 21, Application US/08592936B  
Patent No. 5783393

GENERAL INFORMATION:

APPLICANT: Kellogg, Jill A.

APPLICANT: Bestwick, Richard R.

TITLE OF INVENTION: PLANT TISSUE / STAGE SPECIFIC PROMOTERS FOR

TITLE OF INVENTION: REGULATED EXPRESSION OF TRANSGENES IN PLANTS

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM PC compatible

OPERATING SYSTEM: DOS



SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,936B  
FILING DATE: 29-JAN-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Susan T.  
REGISTRATION NUMBER: 38,443  
REFERENCE/DOCKET NUMBER: 4257-0012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 324-0880  
TELEFAX: (650) 324-0960  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
US-08-592-936B-21

Query Match 17.5%; Score 58; DB 1; Length 342;  
Best Local Similarity 34.5%; Pred. No. 4.3; 21; Indels 10; Gaps 3;  
Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;

3 NFIAGKRYVYIINDGMKKECHGAGNLEFQ---GIRSEQINRLSTRNPCNITSRVYG 56  
Db 236 NMLEBDASVLPFSDDKNTKIKHAKNSLAFDGLKVL-SKINGLDLRN-----NRIYV 287

RESULT 8  
US-09-111-573-21  
Sequence 21, Application US/09111573  
Patent No. 5929302  
GENERAL INFORMATION:  
APPLICANT: Kellogg, Jill A.  
APPLICANT: Bestwick, Richard K.  
TITLE OF INVENTION: PLANT TISSUE / STAGE SPECIFIC PROMOTERS FOR  
TITLE OF INVENTION: REGULATED EXPRESSION OF TRANSGENES IN PLANTS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Denlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/111,573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,936  
FILING DATE: 29-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Susan T.  
REGISTRATION NUMBER: 38,443  
REFERENCE/DOCKET NUMBER: 4257-0012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 324-0880  
TELEFAX: (650) 324-0960  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
US-09-111-573-21

Query Match 17.5%; Score 58; DB 2; Length 342;  
Best Local Similarity 34.5%; Pred. No. 4.3; 21; Indels 10; Gaps 3;  
Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;

3 NFIAGKRYVYIINDGMKKECHGAGNLEFQ---GIRSEQINRLSTRNPCNITSRVYG 56  
Db 236 NMLEBDASVLPFSDDKNTKIKHAKNSLAFDGLKVL-SKINGLDLRN-----NRIYV 287

RESULT 9  
US-09-220-528-64  
Sequence 64, Application US/09220528A  
Patent No. 6284540  
GENERAL INFORMATION:  
APPLICANT: Milbrandt, Jeffrey D.  
APPLICANT: Balch, Robert H.  
TITLE OF INVENTION: Artemin, A No. 6284540e1 Neurotrophic Factor  
FILE REFERENCE: 6029-7998  
CURRENT APPLICATION NUMBER: US/09/220,528A  
CURRENT FILING DATE: 1998-12-24  
EARLIER APPLICATION NUMBER: 09/218,698  
EARLIER FILING DATE: 1998-12-22  
EARLIER APPLICATION NUMBER: 60/108,148  
EARLIER FILING DATE: 1998-11-12  
EARLIER APPLICATION NUMBER: 09/163,283  
EARLIER FILING DATE: 1998-09-29  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 64  
LENGTH: 397  
TYPE: PRT  
ORGANISM: Murine  
US-09-220-528-64

Query Match 17.5%; Score 58; DB 4; Length 397;  
Best Local Similarity 44.4%; Pred. No. 5.2;  
Matches 12; Conservative 4; Mismatches 7; Indels 4; Gaps 1;

22 CHGAGNLEFQIGIRSEQINRLSTRNPC 48  
Db 315 CRGSGNLDQ---ECEQLERSFQNPC 337

RESULT 10  
US-08-533-669A-18  
Sequence 18, Application US/08533669A  
Patent No. 5834592  
GENERAL INFORMATION:  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/533,669A  
FILING DATE: 22-SEP-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-533-669A-18

Query Match 17.3%; Score 57.5; DB 2; Length 732;  
Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNCTITSRYVG 56  
Db 269 KKGDKKK---KKIKIKYIDKEELNKTPIWTRNPDITNEBYG 310

RESULT 11  
US-09-307-143-4  
Sequence 4, Application US/09307143  
Patent No. 6335157  
GENERAL INFORMATION:  
APPLICANT: Gonzalez C.  
APPLICANT: Lange, B.  
TITLE OF INVENTION: METHODS BASED ON LOCALIZATION OF HSP90 TO THE  
FILE REFERENCE: 9882-003  
CURRENT APPLICATION NUMBER: US/09/307,143  
CURRENT FILING DATE: 1999-05-07  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-307-143-4

Query Match 17.3%; Score 57.5; DB 4; Length 732;  
Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNCTITSRYVG 56  
Db 269 KKGDKKK---KKIKIKYIDKEELNKTPIWTRNPDITNEBYG 310

RESULT 12  
US-09-183-861-18  
Sequence 18, Application US/09183861  
Patent No. 6365165  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington

COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/183,861  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/022,765  
FILING DATE: 12-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-183-861-18

Query Match 17.3%; Score 57.5; DB 4; Length 732;  
Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNCTITSRYVG 56  
Db 269 KKGDKKK---KKIKIKYIDKEELNKTPIWTRNPDITNEBYG 310

RESULT 13  
US-09-022-765-18  
Sequence 18, Application US/09022765  
Patent No. 6375955  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,765  
FILING DATE: 12-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-022-765-18

Query Match 17.3%; Score 57.5; DB 4; Length 732;  
Best Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLLFPGIRSEQLNR---LSTRNPNCTISRYVG 56  
Db 269 KKDQDKKK---KKKIKKXYIDKEELNKTKEPIWTRNPDITNEYVG 310

RESULT 14  
US-09-045-284A-2  
Sequence 2, Application US/09045284A  
Patent No. 6265192

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107US1

CURRENT APPLICATION NUMBER: US/09/045,284A

CURRENT FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 386

TYPE: PRT

ORGANISM: Homo sapiens

US-09-045-284A-2

Query Match 17.2%; Score 57; DB 4; Length 386;  
Best Local Similarity 41.0%; Pred. No. 7;  
Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 12 YKNDGKKECHGAGNLLFPGIRSEQ-----LNRST 44  
Db 340 YEKVSRLQKACGDAMNLLGYRHVRSQEQORNLIDLST 378

RESULT 15

US-09-190-911-1  
Sequence 1, Application US/09190911  
Patent No. 6365365

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Tangemann, Kirsten

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107CIP

CURRENT APPLICATION NUMBER: US/09/190,911

CURRENT FILING DATE: 1998-11-12

EARLIER APPLICATION NUMBER: 09/045,284

EARLIER FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 386

TYPE: PRT

ORGANISM: H. sapiens

US-09-190-911-1

Query Match 17.2%; Score 57; DB 4; Length 386;  
Best Local Similarity 41.0%; Pred. No. 7;  
Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 12 YKNDGKKECHGAGNLLFPGIRSEQ-----LNRST 44  
Db 340 YEKVSRLQKACGDAMNLLGYRHVRSQEQORNLIDLST 378

Search completed: March 10, 2003, 18:19:23  
Job time : 3.33068 secs



PT plants and for modulating oocyte maturation

PS Claim 3; Page 52; 72pp; English.

CC This is the amino acid sequence of a novel plant steroid receptor  
CC kinase, designated Bin1, which is involved in the pathway for the  
CC synthesis of the plant steroid hormone, brassinolide. 18 New  
CC Arabidopsis dwarf mutants were identified that lacked the ability  
CC to respond to brassinolide, and were named bin mutants. The bin1  
CC mutations were used to map the gene to a small interval on  
CC Arabidopsis chromosome 4. Bin1 was cloned by map-based cloning.  
CC A Bin1 polynucleotide (see AAX07356) was identified within this  
CC interval by sequencing the wild-type and mutant alleles of this  
CC nuclear acid. Overexpression of Bin1 in transgenic plants provides  
CC plants characterized as having enhanced disease resistance,  
CC increased plant yield or vegetative biomass and increased seed  
CC yield. Bin1 expression may also increase resistance to pesticides.  
CC Inhibition of Bin1, e.g. using antisense oligonucleotides (AON), is  
CC used to render plants male-sterile, and to reduce their stature or  
CC yield, e.g. for creating dwarf varieties. Since Bin1 homologues  
CC may be involved in regulation of the menstrual cycle and uterine  
CC function, Bin1, antibodies and AON may be useful as contraceptives,  
CC for improving success of in vitro fertilisation and to prevent  
CC premature labour. Transgenic animals are also provided, and are  
CC models for studying steroid-receptor interactions or can be used  
CC to screen for therapeutic agents.

XX Sequence 1196 AA;

SO Query Match 100.0%; Score 332; DB 20; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 2.2e-37;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 60  
DB 588 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 647

QY 61 PT 62  
DB 648 PT 649

RESULT 2  
AAB67443  
ID AAB67443 standard; Protein; 1196 AA.

AC AAB67443;

XX 15-MAY-2001 (first entry)

DE Amino acid sequence of an Arabidopsis Br1 protein.

XX Disease resistance protein; Xa21; RKK gene; transgenic plant;  
KW Xanthomonas; plant pathogen; Br1 protein; RCH10 protein.

OS Arabidopsis sp.

XX WO200109283-A2.

PN 08-FEB-2001.

XX 28-JUL-2000; 2000WO-US20714.

XX 28-JUL-1999; 99US-036313.

PA (REGC ) UNIV CALIFORNIA.

PI Ronald P, He Z, Chory J, Lamb C, Li J;

XX WPI; 2001-159858/16.

DR N-PSDB; AAF54982.

XX

PT Chimeric plant receptors comprising a polynucleotide encoding a RKK  
PT receptor containing a heterologous extracellular domain and a kinase  
PT domain from a Xa21 polypeptide, useful for modulating plant responses  
PT to pathogens

PS Disclosure; Page 38-42; 47pp; English.

CC The present sequence represents a Br1 protein. The specification  
CC describes chimeric receptors which are used for modulating plant  
CC responses to pathogens. The receptors comprise a heterologous  
CC extracellular domain (e.g. from a Br1 protein or RCH10 protein) and  
CC a kinase domain (e.g. from disease resistance protein Xa21). The Xa21  
CC gene is a member of disease resistance genes referred to as RKK genes.  
CC When Xa21 is present in a transgenic plant, it confers resistance to  
CC Xanthomonas spp.. The plant receptors containing heterologous domains  
CC are useful for modulating plant responses to pathogens including viruses,  
CC bacteria, nematodes, fungi or insects. The nucleic acids can be used to  
CC confer desired traits on essentially any plant.

XX Sequence 1196 AA;

SO Query Match 100.0%; Score 332; DB 22; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 2.2e-37;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 60  
DB 588 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 647

QY 61 PT 62  
DB 648 PT 649

RESULT 3  
AAE19490  
ID AAE19490 standard; Protein; 1196 AA.

AC AAE19490;

XX 31-MAY-2002 (first entry)

DE Brassinosteroid receptor protein encoded by Br1 DNA.

XX Plant receptor protein; disease resistance protein; plant development;  
KW RKK protein; disease resistance; cell free assay; gene therapy;

KW Br1 DNA; Brassinosteroid protein.

XX Unidentified.

OS Unidentified.

XX Key Location/Qualifiers

FT Misc-difference 162 /note="Encoded by TTT"

XX WO200210367-A1.

XX 07-FEB-2002.

XX 28-JUL-2000; 2000WO-US20604.

XX 28-JUL-2000; 2000WO-US20604.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Chory J, Lamb C, He Z;

XX WPI; 2002-227084/28.

DR N-PSDB; AAD30800.

XX Heterologous polynucleotide encoding chimeric plant receptors for  
PT controlling plant development and disease resistance, has leucine-rich  
PT repeat domain, transmembrane domain, and cytoplasmic protein kinase  
PT domain

XX Discloure; Page 49-50; 54pp; English.  
XX  
CC The present invention relates to heterologous nucleic acid molecules  
CC encoding chimeric plant receptor proteins comprising a leucine-rich  
CC repeat (LRR) domain, a transmembrane domain and a kinase domain from  
CC cytoplasmic RRR (disease resistance proteins) proteins such as Xa21.  
CC The nucleic acid sequences are useful for identifying ligands for  
CC receptor or receptor-like kinase. The chimeric receptors are useful  
CC for controlling plant development and/or disease resistance. They are  
CC used in cell free assay useful for determining the ability of a test  
CC compound to bind to or modulate the activity or expression of the  
CC receptor. Sequences of the invention are also used in gene therapy.  
CC The present sequence is brassinosteroid receptor protein encoded by  
CC btl DNA. This sequence is used in the invention.  
XX  
SQ Sequence 1196 AA;  
XX  
Query Match 100.0%; Score 332; DB 23; Length 1196;  
XX Best Local Similarity 100.0%; Pred. No. 2.2e-37;  
XX Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
DY 1 AANFIAGKRYVYIKNDGKKECHGAGNILEFQGISRQNLRLSTRNPNCTTSRYGGHTS 60  
DB 588 AANFIAGKRYVYIKNDGKKECHGAGNILEFQGISRQNLRLSTRNPNCTTSRYGGHTS 647  
XX  
DY 61 PT 62  
DB 648 PT 649  
XX  
RESULT 4  
ABP00429  
ID ABP00429 standard; Protein; 93 AA.  
XX  
AC ABP00429;  
XX  
DT 24-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:840.  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
PN WO200192523-A2.  
XX  
PD 06-DEC-2001.  
XX  
PF 29-MAY-2001; 2001WO-US10836.  
XX  
PR 30-MAY-2000; 2000US-206132P.  
XX  
PR 29-AUG-2000; 2000US-228716P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkete RA, Leach MD;  
XX  
DR WPI; 2002-106308/14.  
XX  
DR N-PSDB; ABL16181.  
XX  
PT Novel human polypeptides and polynucleotides useful for diagnosing,  
XX hyperproliferative disorders and autoimmune disorders  
XX  
PS Discloure; SEQ ID 840; 1037pp; English.

XX  
CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABLN15762 to ABLN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 93 AA;  
XX  
Query Match 43.8%; Score 145.5; DB 23; Length 93;  
XX Best Local Similarity 51.0%; Pred. No. 1.6e-12;  
XX Matches 26; Conservative 11; Mismatches 13; Indels 1; Gaps 1;  
XX  
DY 5 IAGKRYVYIKNDGKKECHGAGNILEFQGISRQNLRLSTRNPNCTTSRY 55  
DB 44 VIGRPYVYIKNDGKKECHGAGNILEFQGISRQNLRLSTRNPNCTTSRY 93  
XX  
RESULT 5  
AAG79244  
ID AAG79244 standard; Protein; 1121 AA.  
XX  
AC AAG79244;  
XX  
DT 03-JAN-2002 (first entry)  
XX  
DE Amino acid sequence of a gene concerning brassinosteroid sensitivity.  
XX  
KW OSEB1; brassinosteroid sensitivity; d61 locus; rice;  
KW internode elongation; internode cell; lamina joint.  
XX  
OS Oryza sativa.  
XX  
PN WO200173036-A1.  
XX  
PD 04-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-JP02770.  
XX  
PR 31-MAR-2000; 2000JP-0101276.  
XX  
PA (NAG-) NAT INST AGROBIOLOGICAL SCI.  
XX  
PI Tanaka H, Kayano T, Matsunaka M;  
XX  
DR WPI; 2001-616505/71.  
XX  
DR N-PSDB; AAI65842.  
XX  
PT Gene relating to brassinosteroid-sensitivity of plants, useful in  
XX controlling growth and development of transformants including rice to  
XX improve harvest and crop yield for animal feed or dwarfism to enhance  
XX ornamental effect

PS Claim 1; Page 72-80; 87pp; Japanese.

CC The present sequence is encoded by a gene, designated OsBR11, which  
 CC enhances the brassinosteroid sensitivity of plants. The OsBR11 gene is  
 CC located to the del1 locus. The gene is involved in the growth and  
 CC development of rice, for example, internode elongation via the induction  
 CC of the elongation of internode cells and bending of lamina joints.  
 CC The OsBR11 gene is useful in controlling growth and development of  
 CC transformants. Transformant rice plants can be obtained to improve  
 CC harvest and crop yield from providing more animal feed in agriculture,  
 CC or plants with dwarfism can be produced to enhance ornamental effect  
 CC and added value in horticulture and related industries.

CC Sequence 1121 AA;

Query Match 42.3%; Score 140.5; DB 22; Length 1121;  
 Best Local Similarity 46.6%; Pred. No. 1.9e-10;  
 Matches 27; Conservative 11; Mismatches 19; Indels 1; Gaps 1;

5 IAGRRYYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTSRVYGHTSPT 62  
 520 IVGRPYVYLRNDELSECRGKSLFETSIKRPDLISMPKQCNFT-RMYGSTEXT 576

RESULT 6

ABB91366 standard; Protein; 1166 AA.

ABB91366;

31-MAY-2002 (first entry)

Herbicideally active polypeptide SEQ ID NO 577.

Herbicideal; plant; agriculture; herbicide.

Arabidopsis thaliana.

WO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB) BAYER AG.

Tietjen K, Weidler M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicideally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -

Claim 5; SEQ ID NO 577; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicideally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.

Sequence 1166 AA;

Query Match 38.6%; Score 128; DB 23; Length 1166;  
 Best Local Similarity 43.1%; Pred. No. 1.1e-08;

Matches 25; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

5 IAGRRYYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTSRVYGHTSPT 62  
 579 VSGKQFAFVRNEG-GTDCRAGAGLVEFEGIRAEKLERLPHVHSCPAT-RVYSGMTWT 634

RESULT 7

ABB92294 standard; Protein; 1164 AA.

ABB92294;

31-MAY-2002 (first entry)

Herbicideally active polypeptide SEQ ID NO 1505.

Herbicideal; plant; agriculture; herbicide.

Arabidopsis thaliana.

WO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB) BAYER AG.

Tietjen K, Weidler M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicideally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -

Claim 5; SEQ ID NO 1505; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicideally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.

Sequence 1164 AA;

Query Match 34.9%; Score 116; DB 23; Length 1164;  
 Best Local Similarity 40.0%; Pred. No. 5.6e-07;  
 Matches 22; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

5 IAGRRYYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTSRVYGHT 59  
 579 VSGKQFAFVRNEG-GTDCRAGAGLVEFEGIRAEKLERLPHVHSCPAT-RVYSGMT 631

RESULT 8

ABB25323 standard; Protein; 919 AA.

ABB25323;

27-NOV-2000 (first entry)

Eucalyptus grandis cell signalling involved protein SEQ ID NO:642.



[illegible][illegible]

**(FARB) BAYER AG.**

Tietjen K., Weidler M,  
WPI; 2002-269010/31.  
**X** Identifying plant target proteins for herbicidally active compounds,  
**P** comprising aligning and comparing nucleic acid or amino acid sequences  
**R** from plant with nucleic acid or amino acid sequences from non-plant  
**SQ** organisms -  
**D** Claim 5; SEQ ID NO 903; 261pp + sequence listing; English.

The invention relates to identifying target proteins  
comprising aligning and comparing nucleic acid or amino acid sequences from plant  
with nucleic acid or amino acid sequences from non-plant organisms using  
suitable search parameters, where plant sequences having an E-value  
greater by a factor of 3 than the E-value of most similar non-plant  
sequences are selected. The polypeptides or nucleic acids encoding them  
are useful for identifying modulators. The identified modulators are  
useful as herbicides.

Sequence 1143 AA;  
Query Match 28.5%; Score 94.5; DB 23; Length 1143;  
Best Local Similarity 33.9%; Pred. No. 0.00058;  
Matches 19; Conservative 13; Mismatches 21; Indels 3; Gaps

2 ANFIGKKYVVIKNDGMKECHGAGNLFEGAIRSEQLNRLLSTRNPCTISRHYGG 57  
Db 549 SGLSGNTMAVRNVG--NSCKGVGLVERSGIRPERLIGPELKSCDPT-RMYSG 601

**RESULT 11**  
**AAB18199**  
ID AAB18199 standard; Protein: 1436 AA.  
AAB18199;  
AC AAB18199;  
XX  
DT 07-NOV-2000 (first entry)  
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:56.  
XX  
KW Plasmodium falciparum; Chromosome 2; human malaria parasite; vaccine;  
antimalarial; malaria; protozoacide; infection; insecticide.  
XX  
QC Plasmodium falciparum.  
P W0200025728-AZ.  
PD 11-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-US26796.  
PR 05-NOV-1998; 98US-0107131.  
PA (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCI D.  
PA (GAR/) GARDNER M.  
PA (VENTER/) VENTER J C.  
XX  
PI Hoffman S, Carucci D, Gardner M, Venter JC;  
X X  
DR WPI; 2000-365347/31.  
XX  
PT Proteins encoded by chromosome 2 of the human malarial parasite,  
plasmodium falciparum, useful as antimalarial vaccines and in the  
diagnosis of P.falciparum infection -  
XX  
PS Disclosure; Page 133-137; 577pp; English.  
CG The present invention describes proteins and their fragments (I) encoded

[illegible]

```

XX RESULT 12
XX AAB96183
ID AAB96183 standard; Protein; 916 AA.
XX
XX AAB96183;
AC
XX
XX 29-OCT-2001 (first entry)
DT
XX
DE Putative P. abyssi protein #3.
XX
XX Hyperthermophilic archaeon; hyperthermophilic protein.
XX
XX Pyrococcus abyssi.
OS
XX
XX FR2792651-A1.
PN
XX
XX 27-OCT-2000.
PD
XX
XX 21-APR-1999; 99FR-0005034.
PF
XX
XX 21-APR-1999; 99FR-0005034.
PR
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
PA
XX
XX Forterre P, Thierry JC, Prieux D, Dietrich J, Iecompte O;
PI Querrellou J, Weissenbach J, Saurin W, Hellig R;
PI
XX
XX WPI; 2001-126236/14.
DR
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
PT
XX
XX Claim 7; Pages 820-823; 1657bp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF6431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC

```

CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade.  
 CC Note: This patent is in the same patent family as WO2000065062, which  
 CC contains additional sequences as shown in AAB9132-AAB9143,  
 CC AAH75903-AAH75920 and AAG66436.  
 CC  
 SQ Sequence 916 AA;  
 Query Match 19.0%; Score 63; DB 22; Length 916;  
 Best Local Similarity 39.1%; Pred. No. 12;  
 Matches 18; Conservative 5; Mismatches 9; Indels 14; Gaps 3;  
 QY 10 YVYIKDGMKKE-----CHGAGNLL-----EFGG--IRSEQLNR 41  
 DB 821 YVLAGTEGAMKERTPGTCHGAGVLSRKAAATRGVGRDRINQELLNR 866  
 ID AAE12580 standard; Protein; 466 AA.  
 AC AAE12580;  
 XX  
 DT 03-JAN-2002 (first entry)  
 XX  
 DE Pseudomonas syringae pv. tomato (Pto) DC3000 EEL ORF1 encoded protein.  
 XX  
 KW Conserved Effector Loci; CEL, cytotaxic; antibacterial; gene therapy;  
 KW Exchangeable Effector Loci; EEL, disease resistance; transgenic plant;  
 KW eukaryotic cell death; cancer.  
 XX  
 OS Pseudomonas syringae.  
 XX  
 PN WO200175066-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 03-APR-2001; 2001WO-US10698.  
 XX  
 PR 03-APR-2000; 2000US-194160P.  
 PR 11-AUG-2000; 2000US-224604P.  
 PR 17-NOV-2000; 2000US-249548P.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 PA (UINE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.  
 PA (UINE-) UNIV NEBRASKA.  
 XX  
 PL Colimer A, Alfano JR, Charkowski AO;  
 DR MPI: 2001-639361/73.  
 DR N-PSDB; AAD20415.  
 XX  
 PT New nucleic acid molecules encoding proteins or polypeptides of  
 PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci  
 PT genomic sequences, for imparting disease resistance to plants  
 XX  
 PS Claim 8; Page 32-33; 217pp; English.  
 CC  
 CC The invention relates to an isolated nucleic acid molecule comprising a  
 CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas  
 CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)  
 CC genomic sequences. CEL and EEL DNA are useful for imparting disease  
 CC resistance to a plant, by transforming a plant cell with the nucleic acid  
 CC and regenerating a transgenic plant from the transformed plant cell,  
 CC where the transgenic plant expresses a heterologous DNA molecule under  
 CC conditions effective to impart disease resistance, or by treating a plant  
 CC with an isolated protein or polypeptide, by applying the protein or  
 CC polypeptide in an isolated form or by applying a non-pathogenic bacteria  
 CC which secretes the protein or polypeptide, under conditions effective to  
 CC impart disease resistance to the treated plant. CEL and EEL proteins

CC are useful for causing eukaryotic cell death, by introducing a cytotoxic  
 CC Pseudomonas protein into a eukaryotic cell under conditions effective to  
 CC cause cell death. CEL and EEL proteins are also useful for treating a  
 CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into  
 CC cancer cells of a patient under conditions effective to cause death of  
 CC cancer cells, and thus treating the cancerous condition. The method  
 CC further involves administering a targeted DNA delivery system  
 CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,  
 CC to the patient, where the targeted DNA delivery system delivers the  
 CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein  
 CC is expressed in the cancer cells. The present sequence is  
 CC Pseudomonas syringae pv. tomato (Pto) DC3000 EEL ORF1 protein.  
 CC  
 SQ Sequence 466 AA;  
 Query Match 18.2%; Score 60.5; DB 22; Length 466;  
 Best Local Similarity 26.4%; Pred. No. 11;  
 Matches 19; Conservative 11; Mismatches 31; Indels 11; Gaps 2;  
 QY 1 AANFAGKRYVYIKNDG-----KKECHGAGNLLFPGITSEQLN--RLSTRPCN 49  
 DB 387 AMNYVAAEKIRLSKPEKRVAFVGAHTATSCDGVGLAEHGVSLVIDGLKSRATVD 446  
 QY 50 ITSRYVGGTSP 61  
 DB 447 INVKNYGGKLN 458  
 ID ABB63435 standard; Protein; 725 AA.  
 AC ABB63435;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 17097.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PT Venter JC, Adams M, Li PWD, Myers EW;  
 DR MPI: 2001-656860/75.  
 DR N-PSDB; ABL07538.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions  
 XX  
 PS Disclosure; SEQ ID NO 17097; 21pp + Sequence Listing; English.  
 CC  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB57737-AB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC

XX Sequence 725 AA;

Query Match 18.2%; Score 60.5; DB 22; Length 725;  
 Best Local Similarity 38.0%; Pred. No. 20;  
 Matches 19; Conservative 5; Mismatches 23; Indels 3; Gaps 1;

QY 2 ANFIAGKRYIYIKNDGKKECHGAGNLEFGIRSEQINRLSTRNPCNT 51

Db 642 ARFKQGRDYRYEYDNGINP--GAFNPLOQLQELRKVTLARLTCNSDRLL 688

RESULT 15

ABB71725

ID ABB71725 standard; Protein; 1341 AA.

XX ABB71725;

DI 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 41967.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL15828.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -

PS Disclosure; SEQ ID NO 41967; 21pp + Sequence listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC

XX Sequence 1341 AA;

Query Match 18.2%; Score 60.5; DB 22; Length 1341;

Best Local Similarity 40.5%; Pred. No. 44; Mismatches 12; Indels 9; Gaps 1;

QY 7 GKRYVYIKNDGKKECHGAGNLEFGIRSEQINRLSTRNPC 48

Db 676 GKRFYIENTIMKLPQHGIVNLL-----LHRLSYPLC 708

Search completed: March 10, 2003, 18:03:34  
 Job time: 4.35135 secs